Why Do Companies Voluntarily Disclose? A Structural Equations Modelling Approach

Stephen Rae

Submitted for the degree of Doctor of Philosophy

Heriot-Watt University

School of Management and Languages

May 2016

The copyright in this thesis is owned by the author. Any quotation from the thesis or use of any of the information contained in it must acknowledge this thesis as the source of the quotation or information.

Abstract

Many papers have attempted to identify why managers voluntarily disclose information about their companies. This is commonly analysed with regression methods, but these often do not account for endogeneity among common explanatory variables.

This study aims to resolve the endogeneity. Data is gathered on various characteristics of 1,436 UK-listed companies. The primary models of interest are made using Structural Equation Modelling (SEM). This technique allows for correlations and causal relationships between explanatory variables, potentially solving the endogeneity problems common to regression techniques. Regression models are made as a basis for comparison to both existing literature and the SEM.

Analysis of SEMs indicates that Signalling Theory is the most consistently supported explanation of disclosure across all data types used. In addition, several of the best fitting models are those for this theory. As Signalling Theory is the best explanation of disclosure tested, managers are primarily providing voluntary disclosure of information in order to demonstrate the company's recent financial performance (and by extension, the managers' capabilities) in order to attract more investment at a lower required rate of return. Results also indicate some power to Legitimacy Theory, suggesting that disclosure is to some extent provided in order to improve the company's image.

Acknowledgements

I gratefully acknowledge the aid, ideas, and support provided by my supervisory team, Professor Claire Marston and Doctor Santhosh Abraham.

I must further thank those who provided less direct advice, help with other matters that arose during the project, or just served as a sounding board for ideas. There are too many of you to list here, but you know who you are and your support is deeply appreciated.

ACADEMIC REGISTRY

Research Thesis Submission

Name:	Stephen Rae		
School/PGI:	School of Management and Languages		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought (Award **and** Subject area)	Doctor of Philosophy Accounting and Finance

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate:		Date:	

Submission

Submitted By *(name in capitals):*	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in the Student Service Centre (SSC)

Received in the SSC by *(name in capitals):*			
Method of Submission *(Handed in to SSC; posted through internal/external mail):*			
E-thesis Submitted (mandatory for final theses)			
Signature:		Date:	

Contents

Chapter 1: Introduction

1.1: Purpose

At the simplest level, the overall research question here is one asked many times before: What drives corporate disclosure?

Many disclosure studies name a single paper as the starting point for the line of research, that of Singhvi and Desai (1971). Other researchers would later build on the paper with a combination of criticism and expansion to provide a broader and more thorough view of what drives companies to disclose information.

Singhvi and Desai (1971) initially investigated correlations between disclosure and several variables, and later in the paper create a multivariate regression model using all of the variables together to identify a formula for disclosure. The major findings of the process were that a company's size has a significant positive effect on its disclosure and that listed companies disclose far more than unlisted companies. The other variables tested had much smaller or insignificant effects.

In the decades since, the basic idea of Singhvi and Desai's (1971) work has been used many times with varied samples. Despite the range of changes to the research context, the basic findings are rarely much different. Company size has a consistently positive effect on disclosure and, although it has generally changed from listed versus unlisted companies to singly-listed versus multiply-listed companies, listing status has a consistently powerful effect on disclosure. It is possible that Singhvi and Desai (1971) found the major causes of disclosure immediately, that size and listing are the dominant causes of disclosure. However, the question is still asked frequently, indicating a belief that there is more to be learned.

The uniformity of results could stem from a number of sources, but many possibilities have been eliminated over time as researchers examined different samples. Papers in the area were initially US focused, but soon turned to the UK and similar accounting systems without notable change. Research in Continental European systems occurred and again identified size and listing as important. Some researchers have tried multiple countries either to increase their samples or as a comparison, but again with little change. Country cannot be considered the cause of the results. Similarly, those papers that have tested specific industrial sectors with the same results suggest that this is not the cause. The long period of time over which disclosure studies have taken place suggests that timing is additionally not the cause of uniform results.

With a range of national and industrial contexts reaching the same conclusions, there is one more possible commonality between papers: methods. With few exceptions, almost all papers use a regression approach to the causes of disclosure. Disclosure is measured in some way and used as the dependent variable, explained by a number of others that act as possible causes of disclosure.

There are several problems with the very consistent use of regression. First, there is little triangulation of results. Very few papers use a different method, so the results are not independently verified in this way. A consistent result from a different method would reinforce the existing findings as it is found by another means. A different result may be due to some feature of the alternative method, or could indicate that regressions consistently make a Type I or Type II statistical error. Any suitable alternative method of analysis would solve this problem.

Second, the details of the regressions in many papers highlight another problem. The values of the regression sum of squares, a useful measure of the model's fit, are often low. This may be due to the aforementioned non-independence; the model fits poorly because variable interactions are not taken into account. However, it may also imply that models currently lack some possible causes of disclosure. The low fit is barely discussed in the literature; researchers include the values, but few papers include any discussion on the fact that generally less than half of variance in disclosure is explained by the model. Regression is not providing the full explanation of disclosure, although it is not clear whether this is a methodological problem or because some important variables have not yet been identified and included in models. This problem can again be approached using any suitable alternative method. A good fit from another method would demonstrate that low fit is from regression, while a poor fit would confirm that the cause lies in the variables used and that some explanation of disclosure is not currently included in models.

Third, are the common explanatory variables independent? Explanatory variables are assumed to be independent in OLS regression, yet potential connections between common explanatory variables can be made. For example, the ability to draw upon widespread funding from being publically listed means the company can potentially grow beyond the funding limits it would reach as a private entity. Similarly, a company that has reached its limits will be more likely to list in order to grow further. It is not clear whether listing should be considered a cause of size or if size leads to listing, but there is reason to believe these variables are not independent. However, they are used together in a technique that assumes variables to be independent.

Size and listing status, aside from the arguments above, were discussed in literature almost immediately. Buzby (1975) suggested that there would be overlap and criticised Singhvi and Desai (1971) for using both without further discussion. If they are connected, as expected, then regressions using both will be inaccurate due to the correlation. However, the correlation may be weak enough that the results are not changed heavily by including both.

The use of additional variables has the same problem of possible correlation and non-independence. This may explain why few other consistent results have emerged from regression. A variable that influences both disclosure and another explanatory variable will not be treated properly in regression due to the independence assumption, with the effect of understating its importance as an explanation of disclosure. This could either mean other variables have understated importance or that size and listing are even more powerful than is currently thought.

Further, Ahmed and Courtis (1999) performed a meta-analysis of disclosure research to that date, using many papers together to identify the broad trends in results rather than studying the matter directly. For the most part, the results are as expected from investigating other papers, with size and listing status found to be significant as causes of disclosure. However, the paper includes a test for moderation among variables, finding some signs that the relationships between the causal variables are not as simple as the complete independence required of the regression technique favoured by most in the area.

Unlike the other two, this third problem of non-independence cannot be solved with any suitable alternative. A technique that explicitly allows for explanatory variable non-independence is necessary. There are various options available with this limitation; the independence requirement in regression has led to various methods being developed without this restriction.

The chosen method is structural equation modelling (SEM). SEM is a very general case of the same techniques used in regression, including a relaxation of the requirements that all explanatory variables be independent. Using this method would allow the use of the standard variables without the attendant problems of non-independence, and the underlying similarity of methods should mean that the main change made to the interpretation of results is in the effects of this non-independence.

A few additional traits of SEM are used in the thesis and were considered in the decision to use this method. These are summarised here and discussed in more detail in section 5.4. The ability to use latent variables provides a more rounded view of a given concept. For example, multiple measures of the concept of size are used (similar to Cooke, 1992) to provide a balanced measure of the scale of a company rather than being limited to a single measure that may artificially over- or under-rate a given company. The second is related to the relaxed variable independence assumption: SEM does not simply allow explanatory variables to be related, it can actively be used to model cases where this is happening and find the significance and effect size of such links. This is used in modelling theories to provide supporting evidence for a given explanation of corporate behaviour. More detail on the theory aspect of the thesis is provided in chapter 4.

To summarise, there are three research questions:
RQ1: What combination of company characteristics drives corporate disclosure?
RQ2: Which theory or combination of theories best explains disclosure?
RQ3: Does the use of SEM offer any benefit over regression analysis for the research questions above?

The thesis takes a purely empirical approach to the research questions, using statistical techniques to find generalised answers. Regression, the common method of performing such research in this field, is used here as a basis for comparison to test the use of SEM. The intended contribution to literature from this thesis comes in comes in three parts.

The first aspect is the use of SEM instead of regression. Although related, this is a different methodology and should provide a different perspective on the research question of what drives disclosure. Different results indicate a possible problem with using regressions and suggest that other methods need to be examined as possible means to answer this question. Results consistent with the common regressions, on the other hand, help to triangulate the results by providing the same answer from another form of analysis. The contribution is either a suggestion of how to better answer the question or a confirmation that existing results are reasonable. The use of a different method is the primary contribution, examining an alternative approach that accounts for endogeneity of variables.

The second contribution is in theories of disclosure. By allowing causal links among explanatory variables, SEM enables models not possible under regression, which in turn allows new models of the theories of disclosure. When using regression, theory is tested by examining each explanatory variable's significance and power as an explanation of the dependent variable; the most appropriate theory is determined by which one best explains the observed significances. However, few theories of disclosure are this simple and most will have implications beyond how explanatory variables affect disclosure. As a simple example, a theory may state that every aspect of a given company, including its disclosures, is influenced by its listing status. If a regression found listing status to influence disclosure then this would be considered support for the theory. However, this model could not show whether listing status affects any other variable. A structural model, by contrast, could allow listing status as a cause of all other explanatory variables and provide a more accurate means of determining whether the theory holds. If listing status is found to influence disclosure and no other variable then the evidence does not support the theory. While this is an unrealistically simple example, it highlights that SEM allows a more holistic approach to modelling that can provide additional supporting or rejecting evidence for a given theory.

The third aspect of contribution is related to the second as it relies on the relaxation of the explanatory variable independence assumption. As discussed above and in more detail in section 5.3, there is reason to expect some of the explanatory variables to not be independent. This is translated into including links between variables in the models of chapter 7. Including these links should improve the quality of the models by better representing the underlying reality. This may in turn suggest different relationships between disclosure and some explanatory variables; the effects of non-independence in regression are a distortion of the power of the variables involved, making some appear to be more or less important as determinants of the dependent variable than they should be.

To summarise, the contribution this thesis makes to literature is the use of an alternative methodology. The effects of this are expected to be varied and wide-ranging, although the primary contribution is that SEM can be used to expressly allow endogeneity among explanatory variables.

1.2: Summary

The three questions above are answered with reference to a series of models using real-world data. Data is gathered on a sample of 1436 UK-listed companies, with no size or industry sector limitations. The variables taken from this relate to six characteristics of the companies: Size, multiple listing, financial performance, debt finance levels, sensitivity to public opinion, and earnings volatility. Four of these characteristics are not single variables, but are instead represented by several alternative measurements of the underlying concept. Finally, disclosure is measured with reference to each company's total analyst following, using Lang and Lundholm's (1996b) argument and subsequent finding that analysts are more likely to follow companies that voluntarily release more information than comparable companies.

In addition, five theories explaining disclosure have been selected for study. These are all broader theories developed to explain some aspect of accounting or finance that have later been applied as explanations of disclosure: Lemons Theory (Akerlof, 1970), Signalling Theory (Spence, 1973), Legitimacy Theory (Shocker and Sethi, 1973), Agency Theory (Jensen and Meckling, 1976), and the Political Costs Hypothesis (Watts and Zimmerman, 1978 and 1990). The five are compared against each other to determine relationships. Lemons and Signalling Theories are considered identical in terms of disclosure due to very similar necessary and sufficient conditions, while Legitimacy Theory and the Political Costs Hypothesis are considered identical due to similar predictions. Other than these relations, the theories are all found to be unrelated, allowing all to have some explanatory power over disclosure independently of others. Section 4.7 explains these comparisons in more detail.

Two types of model are used here. The first is the same type of regression analysis used frequently in quantitative research into disclosure, using one measure for each of the six characteristics as explanatory variables and examining the effects each has when disclosure is the dependent variable. These largely support past research, finding company size to be particularly powerful, while multiple listing and sensitivity to public opinion consistently have significant effects in addition. Financial performance and debt finance are each less reliable and volatility is almost never significant. Overall, these regressions favour Legitimacy Theory as the consistently significant sensitivity variable is primarily included to test this theory.

The second type of model is the SEM aspect of the project. In the regressions, theories of disclosure are not compared directly. Researchers examine the implications of their findings to determine which model best explains the observed behaviour, based largely on the significance and magnitude of effect of the variables included. In SEM, models are built specifically to represent each theory and the fit of each compared against each other to determine which theory best fits the observed results. In these, the favoured theory is Signalling Theory. In terms of significant variables, the models generally support the regressions, suggesting that the difference in most supported theory is the result of allowing variables to be correlated.

In answer to the research questions, SEM has obtained a different result to the regressions, suggesting that correlation amongst common disclosure study variables is causing problems in regression. Signalling Theory is the best supported of those tested, meaning managers are attempting to make their company stand out from others and obtain a deservedly better valuation in the market. Among the variables suggested as drivers of disclosure behaviour, company size is of high importance, followed by multiple listing status and financial performance.

1.3: Layout

The remainder of the thesis is structured as follows.

Chapter 2 discusses the context of disclosure. The chapter covers definitions of disclosure and some of the reasons put forward both for and against the practice. Additionally, the chapter includes a discussion of the difference between voluntary and mandatory disclosure, as well as the difficulties that occur in distinguishing between the two.

Chapter 3 is the literature review for the thesis. This mostly discusses literature on the determinants of disclosure, split into a few groups based on trends identified over time in literature. In addition, a shorter section covers reasons given against disclosure. SEM literature is discussed along with the method in Chapter 5.

Chapter 4 covers the theories of disclosure. Five theories are emphasised here and each has a subsection discussing its origin and how it has been applied as an explanation of disclosure. Each theory has additionally been examined in a number of disclosure papers, so an additional short literature review aspect is attached to each. Once each theory is explained, a comparison method is used to determine the relationships between each. The final part of Chapter 4 discusses other possible explanations of disclosure.

Chapter 5 explains the data set and the methods of analysis. This includes a discussion of where the data was sourced and how the sample was formed. The next part covers each variable’s inclusion in the study and explains how each is measured. In addition, Chapter 5 contains some basic analysis of the variables, although it does not go beyond investigating correlations and descriptive statistics. The chapter finishes with a discussion of how the regression and SEM methods are applied, including a discussion on the use of SEM.

Statistical analysis is split between two chapters. Chapter 6 covers the basic models, including regression and the results of reconstructing them as SEM. The chapter explains how the models were selected, discusses the obtained results, and investigates the results obtained from the process for possible explanations.

Chapter 7 covers the remaining models, all of which rely on SEM and in most cases model a theory directly. Each model is explained, analysed, applied to the data, and the results investigated for a decision on which theory of those tested best explains the observed data. This is then compared against the results of Chapter 6 to determine whether SEM has provided any new information.

Finally, Chapter 8 concludes by summarising the whole thesis and discussing the answers given to the various research questions laid out in section 1.1 above.

Chapter 2: Context of Disclosure

As the chapter title suggests, the purpose of this chapter is to explain the context of disclosure. There are a few specific matters that need addressing before analysis can begin. It covers some important definitions in the area. The next section, 2.1, explains the background of disclosure, including a definition of the concept of disclosure and some arguments given for and against the concept. Section 2.2 discusses the difference between voluntary and mandatory disclosure and what it means in practice for research.

2.1: Background and definitions

This section will explain the definition of disclosure in this context, covering the reasons for disclosure to some extent an explaining why it is considered useful. Any discussion of these matters links to the related topic of why companies may voluntarily release information.

This section asks a seemingly simple question, but one important to the research: What is meant by disclosure? By definition, disclosure is any release of information made by a company, although this is broad enough to be unhelpful. Research into the voluntary release of an unusual piece of information takes a very different form to research into compliance with any statutory disclosure requirements, yet both are covered by the broad definition of disclosure above. This thesis focuses on voluntary disclosure, so uses of the word "disclosure" used throughout this thesis generally refer to voluntary disclosure unless otherwise stated.

The question of why companies choose to give more information than required is the focus of the thesis. However, beyond the theories examined in chapter 4 there are other explanations given, generally based on the company's directors believing there is some advantage to disclosure. Healy and Palepu (2001), a good overview of disclosure research at the time, explain three of the most common in a wide-ranging literature review, providing some evidence for each as discussed below. Many of these specific reasons for disclosure are consistent with one or more of the broad theories of disclosure discussed in Chapter 4 and cannot be considered unrelated.

The first of the explanations is that disclosure improves stock liquidity. The more information is made freely available about a company, the less likely an investor is to make an unwise trade with someone far more informed. Essentially, the risks of making a suboptimal decision are reduced, making investors more willing to trade in the company's shares. Brown et al (2004) and Brown and Hillegeist (2007) make explicit use of this idea, both papers finding the expected negative link between a measure of information asymmetry and disclosure. As an example of the consistency with theory described above, the reduction of asymmetry makes the company a better investment decision to many potential investors, which is clearly in keeping with Signalling Theory (described in section 4.3) and may be argued to work with other theories discussed in chapter 4.

The second reason for disclosure is more common in the literature: disclosure reduces the cost of capital. One potential mechanism for this idea involves an information risk premium (Elliot and Jacobson, 1994). Under this idea, the market will request a higher return from those companies for which little information is available. This is done to compensate for the risk of some unknown factor harming the firm and, by extension, any investment in it. Essentially, the market assumes there is risk where there is insufficient information to be confident there is not. Linked to the above explanation, Brown et al (2004) argue that investors may demand a higher return in low-disclosure firms in order to compensate for the risk of making a poor trade, further increasing the cost of capital.

For a time, this explanation was widely accepted. A number of papers use it without question (e.g. Gelb and Strawser 2001, Healy et al 1999). More recently some papers have argued against the point and found evidence that calls it into question (e.g. Beyer et al 2010, Cheng et al 2006). The likely causes of this uncertainty can be seen in the earlier papers testing the idea. Botosan (1997) models the effect using empirical data and finds that the cost of capital does fall when companies increase disclosure but the effect is far more noticeable when the analyst following was low, suggesting that analysts have a role in revealing information about the risks of investing in a company. Botosan and Plumlee (2002) found that some disclosures reduce the cost of capital while others have no effect. Either of these effects could cause mixed results in practice. Regardless of the cause, the change in general opinion over time is apparent in two ICAEW statements. In 1999 the organisation (ICAEW, 1999) encouraged companies to disclose risk information using a potential lower cost of capital as an incentive, but later (ICAEW, 2011) stated there is no evidence that disclosure has an effect on the cost of capital.

The third and final reason Healy and Palepu (2001) give for voluntary disclosure is that it increases analyst following. This idea is not discussed in detail here as the argument is used in section 5.2.7 to explain the use of analyst following as a proxy for disclosure.

To summarise, analysts serve as information intermediaries and this is useful to companies for various reasons.

In addition to these potential reasons for disclosure, there are several possible reasons to not disclose. Lundholm and van Winkle (2006) offer a few simple possibilities based on a lack of knowledge. If managers do not know some piece of information, it cannot be disclosed. Similarly, if managers do not know the importance of something, it is unlikely to be considered for disclosure. A third possibility given is that the managers do not care enough to disclose, although this raises questions about their competence and dedication to their role.

A more common argument against disclosure, particularly mandatory disclosure, is based on proprietary information and competitive disadvantages (e.g. Dye, 1986). The basic argument is that a company may be forced to release some information that its competitors can use against it. The damage that can be caused is limited if the competition must also reveal the same information as the company and its competitors are in the same situation. For example, if strategic information is part of the mandatory disclosure package then two competing companies, Company A and Company B, must each release information about their strategy and each can use the other's disclosure against them. For similar companies, this is fair as each has comparable capability to act on the information and affect their competitor. For less similar companies, one may have more potential to harm the competitor than the other; if Company A is much larger and works in multiple lines of business, then it is likely able to take a loss in one area in order to undercut or otherwise outperform Company B. The problem becomes more pronounced when the competition is subject to different rules, for example being located in a country that does not require the same information to be disclosed. In this case, Company A receives detailed information on Company B's plans, but Company B does not receive comparable information on Company A.

2.2: Voluntary and mandatory disclosure

While the thesis refers only to "disclosure" in most cases, an important divide between voluntary and mandatory disclosure exists and is discussed here. As the names suggest, mandatory disclosures are required by the regulators or other authorities to whom a given company is subject, while voluntary disclosure is any information released beyond that required of the company. While this seems like a simple definition of the two, in practice they can be difficult to disentangle.

The line between voluntary and mandatory becomes less clear when a company is subject to multiple authorities, usually as a result of listing in multiple locations. Two different countries may have different listing requirements, including different rules on what information must be released in order to retain listed status. For example, a company is listed in one country that requires risk management information be disclosed as part of the annual report and another that requires information on the board of directors in the annual report. While each country requires only one of these pieces of information, the company's managers are likely to release one report that covers both sets of requirements and to release this in both countries. While tailoring reports to each country's rules would be possible, report users in each would be able to use the additional information released in the other to inform their investment decisions. While no information has been made available that the company was not required to release overall, to investors in each of the listing locations there is something available that was not required of the company. The release of risk information is effectively a voluntary disclosure in the country requiring only director details, and the director information is voluntary in the risk management country, but to the company's managers both were mandatory disclosures.

The distinction blurs further when additional effects are considered. Best practice guides and the disclosure policy of competitors can each create a form of non-statutory mandatory disclosure. Once investors have come to expect a certain level of disclosure or specific items of information to be available, failure to meet the expectations will lead to a negative opinion. Holland (1998) describes this as companies competing for credibility; one that discloses less than comparable firms is seen as less credible than its competition. The result is that a supposedly-voluntary disclosure can become effectively mandatory.

In addition, there are questions of how mandatory and voluntary disclosures interact (Patelli and Prencipe, 2007). The question is whether the two types are complements or substitutes. The substitution argument holds that a given company's managers will reveal a fixed quantity of information in total. If regulation forces some information to be released, this reduces the voluntary disclosure level. As a simple example, a company chooses to release two pieces of information, its strategy and the composition of the board of directors. If regulation is introduced requiring that the company disclose risk information, the managers will still release only two pieces of information. The disclosure total will then become the mandatory risk details plus only one of the board or strategic details. Alternatively, the complement argument holds that the company discloses whatever information its managers consider appropriate rather than working towards a total level of information. In the same example, the addition of a requirement to disclose risk information makes no difference to the voluntary disclosure and the company continues to show the board composition and its strategy in addition to the mandatory risk information.

If the substitution argument describes actual practice, mandatory disclosure may have clear adverse effects on the overall information released. The company managers have chosen to reveal what information they consider most appropriate, but stop providing some of this information if forced to reveal something else. There is only a net gain if the information mandated is more useful than whatever the company stops revealing. This may not be the case for all companies; some may end up giving more useful information overall, while for others the previous disclosure package was optimal for investors and replacing one piece of information with another reduces the overall usefulness of the information released.

If the complement argument instead holds, there is much less of a problem. Any mandatory information is revealed in addition to whatever the company managers have previously decided is appropriate. It is possible, though rare, that the new information is confusing to investors in some way. There is also potential for the mandatory information to be of little use, such as risk information in a particularly low-risk enterprise, but not harmful to investor understanding, in which case the mandatory disclosures are neither useful nor harmful to investors.

The discussion above assumes that managers are revealing information with good intentions, aiming to spread information the market will find relevant, and has correctly identified what information will be useful to this end. Should either of these conditions not hold, the mandatory disclosure may be more useful to investors by releasing information of value to them that managers have not been making available.

The testing performed in later chapters (6 and 7) refers to voluntary disclosure. In practice mandatory disclosure cannot be distinguished from voluntary due to the use of a proxy measure of disclosure in this study, analyst following (explained in section 5.2.7). Mandatory disclosure is required from all companies, making it a baseline level of disclosure, and the proxy should respond only to information released voluntarily beyond that required by disclosure regulations. However, the two are impossible to fully separate. Levels of compliance with otherwise mandatory standards vary among companies as some managers will comply with the letter of the rules and disclose exactly what is stated as required while others comply with the spirit and may discuss matters not directly stated but clearly related. Compliance with mandatory disclosure is a complex issue that cannot be given sufficient treatment here. The assumption throughout the thesis is that mandatory disclosures are enforced and serve as a comparable baseline for all companies, making it comparable to the constant in a regression equation, but this may not be the case in practice. Further, there may be additional influences on disclosure that come from sources other than managers; Ahmed and Courtis (1999) do not observe a significant direct effect on disclosure practices from a company’s auditor, but find this to have a moderating effect that results in better mandatory disclosure from those engaging a larger and more prestigious auditor.

The inability to fully distinguish between voluntary and mandatory disclosure is not unique to proxy measures (either this specific one or in general). The alternative approaches, discussed in section 5.2.7, will encounter many of the same challenges. In part, this is due to the above discussions. Differing compliance levels mean some mandatory disclosure is effectively voluntarily, while intra-industry effects and the publication of guidance documents mean some voluntary disclosure is effectively mandatory. Any measurement of disclosure will encounter challenges with the voluntary/mandatory divide due to these effects. A disclosure index approach can either not score a mandatory item, missing any variation in how this is reported, or attempt a highly subjective approach of only scoring sufficiently detailed reports. The other common approach, using a pre-existing scoring system, may or may not take these matters into account depending on whether the original designer considered them.

There is some literature that suggests finding ways to measure the quality of disclosure rather than the (often implicit) quantity approaches commonly used. Beattie et al (2004) discuss several possibilities, ranging from content analysis approaches to using the common disclosure indices with a weighting system that rewards particularly clear and obvious information. As an alternative, Linsley and Lawrence (2007) discuss the use of a readability index for annual reports, a measure of how clear the report is that should provide a quality measurement by penalising examples where information is unclear. Disclosure quality measurement is a field of study in itself and cannot be treated with the detail it needs in this thesis.

2.3: Conclusion

Disclosure, for the purposes of the thesis, is defined as voluntary releases of information above and beyond that required by regulation.

The information in this chapter is a combination of discussion of voluntary and mandatory disclosure and some arguments for why a company's reports may contain further information. Research in the area has examined not only whether more information than required tends to be disclosed, but also attempted to determine which companies are most likely to do so and test the proposed explanations for why this practice occurs. These matters are discussed in the next chapter.

Chapter 3: Literature

The chapter focuses only on literature on the driving forces behind corporate disclosure. While other lines of research are used in the thesis, such as papers defining and using the theories discussed in chapter 4 and a discussion on the use of SEM in chapter 5, these are placed in the relevant chapters. Research into disclosure has been split into three broad groups based on the purposes and themes of the research.

The first group consists of the early papers in the area. The generally accepted start of research into the determinants of disclosure occurs in the early 1970s, and over the next decade further researchers attempted to expand on the initial work in various ways. These efforts collectively make up the exploration group and are characterised by their exploratory nature. Researchers are generally trying new things at this point, but quickly create what would become standard practices for later researchers. This is termed the exploration group as a reflection of its early nature and is discussed in section 3.1.

The early 1980s saw research in the area reduce in frequency. By contrast, the end of the decade marked the beginning of a resurgence which continued through the 1990s. There does not appear to be any single event that explains the reappearance of such research, but the nature of the research offers a possible explanation: the timing coincides with the spread of accessible computing power while the research tends towards large samples and statistical techniques. Regardless of the explanation, the papers in this group usually take the methods established by the first set and apply them to new, large data sets. This is termed the core group as it generates much of what becomes common knowledge of disclosure and is discussed in section 3.2.

As the 2000s began, the path of research forked, leading to two concurrent groups of papers. One of the new paths is a straightforward extension of the second group. Papers to that point had largely focused on general disclosure in a few countries, particularly the USA. The extension of the group took the now firmly established methods and variables, but adds something unusual each time. A given paper may focus on a country not previously tested or perform an international comparison, examine a very specific form of disclosure, or restrict the companies examined in some way to investigate specific subsets of the market. These papers, which are considered a separate subgroup, are characterised by a combination of a late 1990s or later date and a focus unlike that of earlier papers. A rare few try different methodologies, but the majority apply established techniques to different samples. This is named the expansion group as it takes the core group ideas to new areas. The group is discussed in section 3.3, which is further split into papers looking at different countries and papers using specific forms of disclosure instead of a general measure of all information a company releases. The other path taken by many is a response to new regulation and guidance encouraging new forms of reporting, primarily relating to risk information.

As the man body of the second group reached publication in the early 2000s, a number of changes to the disclosure environment occurred. In the USA, Regulation Fair Disclosure (often shortened to Regulation FD or just FD) was enacted 2000, while the Enron scandal led to the Sarbanes-Oxley Act (SOX) in 2002. Each had some effect on disclosure, FD by making private disclosure to privileged parties largely impossible and SOX by making it easier to sue directors for information they have not revealed, leading to a reasonable expectation of changes in disclosure practices. Elsewhere in the world, the EU officially adopted IFRS as union-wide financial reporting standards in 2005 after several years of laying foundations for this move. In the UK, the Accounting Standards Board at the time was encouraging narrative reporting with an emphasis in risk management (e.g. ASB, 2007a). All of these and similar events led to disclosure papers that examine the effects such changes had on the disclosure environment. Many compared disclosure before and after the relevant change occurred (e.g. pre- and post-Regulation FD) to directly examine changes. This group is named as the response to changes as the papers always examine disclosure practices following some change to the overall environment and discussed in section 3.4. It is not discussed in detail, however, as the papers tend to focus more on the changes than the disclosure itself and are of less relevance to this thesis. Papers are generally included if they have some idea that has been used elsewhere in the thesis, such as explaining the use of a particular variable.

3.1: Group 1: Exploration of the concept

A small number of papers are frequently cited by later works as starting points for the research. These three early papers are summarised in table 3.1 below

Table 3.1: Exploration Papers				
Paper	**Sample**	**Dependent Variables**	**Significant Variables**	**Theories mentioned**
Singhvi and Desai 1971	155 listed and unlisted US firms	Disclosure index	Size + Shareholders + Listing + Profit + (2 measures, EM and P/S) Big X Auditor +	Reduction of information asymmetry
Buzby 1975	88 US firms, equally listed and not	Disclosure index	Size +	None, critique of previous paper
Firth 1979	180 UK firms, mix of listed and unlisted	Disclosure index	Size + Listing +	Speculates about costs and benefits

As mentioned earlier, the literature frequently credits Singhvi and Desai's 1971 analysis as the beginning, although it should be noted that this paper explains that it is built on an earlier work. However, accuracy of the claim notwithstanding, Singhvi and Desai's paper is the agreed starting point for the entire line of research. It is a clear and relatively straightforward piece of research that justifies its ideas well enough that the majority of later works in the field build on the paper. For these reasons, it is a good starting point for researching the area of study.

The paper compares six company characteristics against the company's disclosure for a total of seven variables, almost all of which would be used repeatedly in later research. The disclosure is measured with reference to an index that the authors have created and searching for items in the annual reports, making it a quantity of information measure. The explanatory variables are more mixed.

Company size is measured as total assets as recorded in the annual report. It is justified with multiple reasons. First, a larger company is more likely to gather such information in order to make overviews of operations for senior managers who cannot realistically oversee the entire business. Second, larger companies are more likely to raise funding in securities markets, so benefit more from an informed market. Third, the sensitive information argument sometimes used against disclosure (see 2.1 above) is here said to be worse for smaller firms, making them less willing to reveal information. Size is used in nearly all later papers using similar research methods, although the exact measurement varies.

The second explanatory variable is the number of shareholders. Having more shareholders is argued to mean more visibility and the resulting pressure, plus the potential for new disclosure regulations if the company is not seen as giving enough information. Shareholders are stated to be more likely to be concerned with social responsibility than managers, meaning that they want to see such information disclosed. Further, the paper suggests the simple matter of marketability of shares. This variable is one of the few that has not been commonly adopted by later researchers.

The third variable tested in the paper is listing status. The reason given is that listing rules require some information to be released, so higher disclosure is expected among these companies for that reason. While a reasonable item to study, this variable represents a possible self-selection problem in the sample. The unlisted firms in the sample were initially chosen randomly, but the lack of publicly available information meant that the companies in question had to agree to provide information, leading to a possible response bias. For this and similar reasons, later papers have generally not used listing in this sense. Where it is used, the variable tends to be multiple listing in some form.

The fourth variable is named as CPA firm. This essentially asks which financial services company is providing aid with the financial reporting process. The justification given is that a larger, more prestigious accountant is more likely to exert influence to reveal more information. The larger firms have their own reputations to uphold and making detailed accounts of their clients is a possible means to demonstrate their own competence. Later research frequently uses a slight variation on this idea, examining the auditor rather than the accounting firm, but otherwise using the same reasoning. The only other difference is one of changes to the financial services sector over time. Singhvi and Desai (1971) measure the CPA firm as a binary large or small classification, where large means the CPA firm is one of the Big 8. Variables like this have changed a little over time as the Big 8 financial services providers have merged with each other, so later papers using the same concept will refer to the Big 6, Big 5, or Big 4 depending on the timing.

The fifth variable is the rate of return, measured as a ratio of profit to worth. A high ratio will be seen as a sign of good management and encourage managers to disclose more information. Similar ideas have been used in later papers, although the means of measuring the return vary greatly and tend to be named as financial performance measures.

The final variable used is the earnings margin, the ratio of profits to sales. The paper argues that the rate of return shows overall performance while this covers the ability to deal with rising costs and therefore stability. A low ratio makes it easy for the competition to squeeze the firm out with pricing wars, while a high one allows the firm to reveal more about its operations since it will still be hard to pressure them out of the line of business and this reassures the shareholders of stability. While this variable is used in later papers, it usually appears as a financial performance measurement.

These variables are analysed in two different ways. The first method is the simpler of the two. Each variable is split into a few size categories and compared against similar categories for disclosure, giving a general relationship in each case that matches the expectations mentioned in the explanations of each variable.

The second method is a regression analysis, using the six variables to attempt to explain the disclosure behaviour more directly, the method adopted by later papers analysing the situation. These results suggest (at first glance; see Buzby (1975), below) that the size and listing status of the company are the most important of all explanations.

Two problems stand out on closer analysis. First, the regression coefficient R^2 is somewhat low at 0.43. While this indicates some explanatory power, the model is explaining a little under half of the total variation in disclosure observed in practice. However, later research has tended to obtain similarly low values. This is discussed in more detail under 3.6 below.

The other problem comes from further analysis Singhvi and Desai performed. When looking at the power of individual variables, the R^2 contribution of listing status was identified as 0.38 (out of a total of 0.43), making it firmly dominant. Buzby (1975) argues that this strong effect indicates problems in the analysis. While listing status should be expected to have an effect, Buzby holds that the link between disclosure and size should be far more powerful than is demonstrated. To back up this argument, Buzby tests the correlations of firm size and disclosure levels. For listed firms, the correlation is 0.515, while unlisted firms have a lower 0.370. These numbers are high enough to suggest that a stronger relationship would be expected, leading Buzby to speculate as to why Singhvi and Desai (1971) did not find one. The main suggestion given is that size and listing are themselves linked, leading to a situation where the effects one has on disclosure are difficult to identify separately from those of the other. If true, this could explain the difference between Buzby's (1975) expectations and the original paper's results. The assumption of independence among explanatory variables may not be true in this case. However, while the arguments are good, Buzby does not test the ideas at all, leaving the research as a series of logical but untested reasons why Singhvi and Desai (1971) may not have obtained meaningful results.

Firth (1979) uses the questions of independence in a UK-based study based on Singhvi and Desai's (1971) work. Firth argues that the majority of the variables tested are likely to be related to size and cuts three to leave size, listing, and auditor, noting that listing status is likely to be related to size. The use of auditor rather than more general financial service provider is argued using prestige. The auditors themselves have powerful reputations that keep them in the Big 8 group (as it was at the time), so will likely pressure clients to release more information in order to keep that reputation. The client, meanwhile, wants the prestige of passing the strict standards of one of the Big 8's strict rules, so will tend to follow any auditor advice given on disclosure.

Firth's analysis is quite different to the paper he is building on, not performing any regression. Instead, Firth performs a correlation test of size and disclosure, similar to Buzby (1975). The other two variables are different in nature as listing status and Big 8 auditor usage are binary variables. To examine the effects of these on performance, means differences tests are used. The sample consists of 40 listed firms and 40 unlisted firms that are each as similar to one of the listed ones as possible. A significant difference in disclosure is found when testing listed versus unlisted; as expected, listed firms disclose more. The Big 8/other auditor grouping, by contrast, displays no significant difference. Firth does not speculate as to why this is the case.

At this point, the research into disclosure has been limited in scope and a conclusion is difficult to draw, largely due to the small number of papers involved. Two possible methods have been explored: Buzby (1975) and Firth (1979) investigate correlations among variables to determine how disclosure reacts to changes in other variables, while Singhvi and Desai (1971) explore a multivariate regression model of disclosure. In terms of findings, listing status and the size of the company are identifiable as the main determinants of disclosure. Later papers build on both the method and the variables examined, as discussed below.

3.2: Group 2: The Core Research

As the line of research picks up into the 1990s, certain standard practices emerge. Most papers take a sample within a single country. While exceptions exist, the focus at this point is largely on finding out what is happening in a single location, and in some cases within a single industry. A common set of variables can be drawn up, largely based on Singhvi and Desai's (1971) set of variables. Size is in almost all papers, included as a control variable even where it is not of interest to the researcher, and is almost always positive with disclosure. One of the few exceptions comes from Malone et al (1993). This research is typical of the area aside from a focus purely on oil and gas companies, a subset for which size does not link to disclosure practices. The three significant variables here are stock exchange listing, number of shareholders, and the use of debt finance.

Some form of listing variable is common, although the difficulties of getting information about unlisted firms means that over time multiple listing becomes more common than anything involving unlisted firms. Whichever is used, a generally positive relationship is found. Some measure of the firm's profitability is commonly included, although it is insignificant about as often as significant, and even among significant results it is positive as often as negative. CPA firm is largely replaced with an auditor size measure following Firth (1979), with a generally positive result. Various stakeholder measures, including Singhvi and Desai's (1971) number of shareholders among others, are common but somewhat inconsistent in their effects on disclosure.

In addition to these, a few other common variables have emerged. The use of debt finance in the firm has become a common variable, first appearing among identified papers with Malone et al's (1993) study and appearing commonly from that point, although its results are inconsistent between papers (note that, as explained in Chapter 5, the predictions with this variable suggest either positive or negative effects depending on the arguments used). Corporate governance measurements have become more common over time, again first appearing in Malone et al (1993) among the identified papers, but unlike debt finance there is a delay before it appears in another identified paper (Bushman et al, 2004). It is found to be generally positive with disclosure, suggesting that information release is seen as responsible management.

Industry membership has become a common control variable as researchers have identified that different industries have different practices, first seen in the identified papers with Clarkson et al (1994). Similar variables are common for the next decade, but become less frequent after Field et al (2005). As a categorical variable it is difficult to summarise its effects on disclosure beyond observing that industries the researcher expects to disclose more generally do so.

The final common variable relates to the volatility of the company's earnings, although this has inconsistent effects overall. Among identified papers this is first seen in Clarkson et al (1994), but is not used again for a number of years (Gelb and Strawser, 2001), after which it appears with some frequency.

Although research on the matter is less common, it is thought that companies could increase disclosure in advance of any market activity they perform such as share issues in order to make their stock more attractive. This is tested with indicator variables of market activity within the next year or two and is generally supported, but Frankel et al (1995) offers a strong counterargument. Rather than altering disclosure levels for immediate gains, Frankel argues that changing the disclosure practice is a rare and possibly unpopular event. Instead, a time series analysis suggests that disclosure comes from a disclosure policy rather than the usual variables, although the policy will be set in accordance with company characteristics. This ultimately changes very little compared to past research as the same variables seen in other papers have the same overall effects on disclosure. The only difference is that these variables are used to set the disclosure policy and this is a slow process, rather than the usual assumption that disclosure can vary year-on-year in response to changes to the variables.

Other, less common variables appear on rare occasions and tend to be core aspects of the research in question. Zarzeski (1996) offers a good example of this. While the paper ultimately becomes an example of a disclosure regression, its stated purpose is investigating the harmonisation of accounting practices. This means most of the "normal" variables are included for control purposes only and the variables of interest are four cultural factors identified by Hofstede (1984) and later translated into likely effects on accounting systems (Gray 1988).

A summary of papers in this group is presented in table 3.2 below

Table 3.2: Core Research				
Paper	**Sample**	**Dependent Variables**	**Significant Variables**	**Theories mentioned**
Cooke 1989a	90 Sweden-listed firms	Disclosure index	Size + Mult. Listing + Foreign operations **+**	Focus on importance of listing status
Patten 1992	21 US-listed oil firms	Content analysis	Size + Involvement in crisis +	Explicitly names Legitimacy Theory
Malone, Fries and Jones 1993	125 oil and gas firms	Disclosure index	Stakeholders + Listing + Debt finance +	Looking for differences in disclosure levels
Clarkson, Kao, Richardson 1994	905 Canadian firm-years	Existence of a voluntary forecast	Size + Market action + Volatility + Industry effects	Importance of financial market considerations
Frankel, McNichols, Wilson 1995	1880 US-listed firms	Probability and/or existence of a voluntary forecast	Debt finance + Market action +	Importance of finance needs to disclosure decisions
Ahmed and Courtis 1999	29 papers	Meta-analysis; other papers' dependents	Size + Listing **+** Debt + Big X auditor +	Meta-analysis
Botosan and Harris 2000	107 US firms with multiple segments	Existence of quarterly reports	Industry effects	Asymmetry; reasons for quarterly reports
Gelb and Strawser 2001	233 US listed firm-years	AIMR rankings	CSR rating + Performance +	Indirectly supports Legitimacy

Cooke (1989a) is one of the early papers in this second group of papers. The research contains an unusual listing variable, which is the focus of the research. In this case, there are three listing categories: unlisted, listed, and multiple listed, which is included because it is expected to increase disclosure. The sample is drawn from the Stockholm Stock Exchange which, at the time, had an unusually large number of multinational companies listed. This non-US focus is, as table 3.2 suggests, very unusual for the time.

While other variables are mentioned, the final models here have only two: listing and size. The results agree more with Singhvi and Desai (1971) than Buzby (1975), finding that listing status is very powerful as a determinant of disclosure. In addition, this is a more powerful test, having R^2 0.60, far above the 0.43 of Singhvi and Desai (1971) and, notably, meaning that the model here explains over half of the disclosure variation in the sample.

Clarkson et al (1994) offer a number of notable points. Their paper is limited to examining the specific disclosure of a forecast for future results. In a sample from the Toronto Stock Exchange, the firms found to make forecasts most often tend to be larger and found in certain industries more than others. This latter point is not given much attention in the paper, but it is worth noting that the industries least likely to forecast tend towards extraction (e.g. mining, oil and gas), while those most likely to forecast are service industries (management, financial services). Generally speaking, it appears that the forecasters are more intangible in nature, although the paper does explain that the resource extraction industries tend to have fluctuating prices for their outputs that make accurate forecasts very difficult.

The research further examines the tone of forecasts, dividing them into positive, neutral, and negative (indicating, respectively, increased income, similar to previous results, and lower income). Just over two-thirds of all forecast disclosures are positive in nature, with the other two splitting the remaining numbers roughly equally. Interestingly, when the tone of past forecasts is compared to the actual results, there seems to be an upwards bias. Approximately a third of positive forecasts are incorrect, while the majority of neutral forecasts are later proven to be negative results.

Negative forecasting may appear harmful to the company as it will lead to lowered expectations in the market and requires managers to admit to failure. However, there is a point early on in the paper that, with any information release, companies are attempting to balance the need for finance with the need to discourage potential competitors. This results in the company disclosing good news for the investors and bad news to suggest to other firms that they should not enter this competitive space, attempting to find the perfect balance of these two aspects. Gigler’s (1994) model of disclosure includes the same concept and ultimately concludes that disclosure must be a fine balance between pleasing investors with promising future outcomes and deterring potential competitors with low expectations. Clarkson et al's (1994) results are an analysis-derived confirmation of the model.

Ahmed and Courtis (1999) offer a useful summary of disclosure research of this form to the time of publication. The primary methodology in use here is a meta-analysis, effectively attempting to merge existing results together in a new, more wide-ranging regression-like summary that will highlight the most significant and powerful of the commonly-used variables. The results of this research are effectively an average of past papers, weighted by their sample sizes. The size and listing status are by far the most important variables across the research to that point. Size is found to be quite powerful but with a large variance and surprisingly low explanatory power, suggesting that other variables may be necessary for a full explanation of disclosure. Listing is less powerful, but explains more variance in disclosure, in keeping with Singhvi and Desai's (1971) results. The third commonly-tested variable that this analysis finds to be significant is the debt finance ratio, which is somewhat unexpected as past research has found this to have inconsistent effects.

As with any methodology, meta-analysis has its own flaws. In using studies from various time periods and international contexts together, there are obvious questions about whether the component research papers can be considered directly comparable. Further, any research method based on large samples will result in potentially interesting and meaningful details about individual cases being glossed over; this is the disadvantage of large sample techniques. With meta-analysis, this becomes even more pronounced as the details of entire studies are passed over. However, the core advantage of large sample methods is that the results can be generalised across a wide range of cases, and meta-analysis similarly amplifies this aspect by effectively using a range of samples.

In addition, Ahmed and Courtis (1999) discuss variable interaction. The paper mentions that some of the common variables in disclosure research may be related to each other. The main aspect mentioned is moderation, in which the relationship between two variables is influenced by a third. The typical example of moderation is a case where X causes Y, but the strength of the effect depends on the value of some additional variable M. The argument here is that, for example, the company's auditor is not significant overall because its effects depend further on where in the world the company is located and the local accounting system. The usual effect of uncontrolled moderation is to reduce the significance of the affected variables. Three possible moderation effects are suggested in the paper and tested, finding some validity to the ideas, albeit with a somewhat low effect in most cases.

It should be noted that the paper does not discuss the related problem of mediation. This is similar to moderation in that the relationship of two variables is influenced by a third, but in a very specific way. Using the same example variables above, a mediation relationship is characterised by X having a causal effect on both Y and the additional variable M, plus M having a causal effect on Y. In effect, the explanatory variable X may affect Y both directly and indirectly through its effects on M. The usual effect of mediation is to overstate the importance of X, as its effects on Y are a total of both its direct effects and the indirect effects through M. This represents a problem for existing research; regression methods do not allow for such interactions, and if they are genuinely occurring then the common research techniques are not entirely appropriate.

Healy and Palepu (2001) offer a second overview, in this case taking the form of a wide-ranging literature review; many aspects of disclosure are covered, but few in great detail. As discussed in chapter 2, the topics covered include the common determinants, managerial incentives, and the actual effects of disclosure on, for example, the company's cost of capital. Some of the major theories of disclosure are mentioned throughout. This is given more discussion later when covering disclosure theories.

Core (2001) offers a critique of the previous paper, however. The main point is that Healy and Palepu (2001) were not critical enough of the research being examined. In all of this, Core (2001) makes an excellent point: Disclosure policy is an assurance of disclosure quality. A firm that commits to disclosing as much information as possible will by definition produce detailed, high-quality information in its disclosures. By contrast, one that has a policy of allowing managers to screen the disclosures will tend to see less information come out if they have any incentive to hide bad news, making the company's disclosures less meaningful.

Core's (2001) stated purpose is to complement Healy and Palepu's (2001) work. The aim is to provide more detail to add to their discussions of voluntary disclosure, adding the depth lacking from the original. The main point of contention is that Healy and Palepu (2001) focus on an argument that states disclosure can only be a good thing as it provides more information, while Core (2001) is more willing to look at the negatives involved – usually direct costs of gathering information or providing incentives. This argument is logically extended to the implication that stable firms are likely to disclose less than growing companies simply because disclosure as a form of information release is useful for attracting new investors, but the costs involved mean that it will be less appealing where the company is stable and has less need for such.

Botosan and Harris (2000) take a different approach to measuring disclosure. In this case, rather than any indication of the overall quantity of disclosure given out, the researchers have looked at the frequency of disclosure. The research question here is about which companies will tend to disclose quarterly instead of the lower frequency normally mandated by regulators. The set of explanatory variables is similarly unusual. While conclusions can be drawn, primarily finding that firms identifying information asymmetry are most likely to use this practice, this is not easily fitted into the wider picture being drawn from more standard approaches. In addition, there is a noted follow-the-leader effect in that industries with a history of quarterly reports seem to encourage those making less frequent disclosures to increase frequency, which suggests that disclosure levels are in part determined by factors external to the disclosing firm.

In closing, this group of papers represents a maturation of the ideas from the first papers discussed in section 3.1. The ideas have spread and been used in a large number of papers. The volume of research allows generalised conclusions to be drawn, leading to at least one meta-analysis of other papers. However, little changes compared to the initial findings. Company size remains a major determinant of disclosure activity. Listing status begins to change form, moving away from listed versus unlisted companies and becoming listed versus multiple-listed companies, but largely retains its significant effect on disclosure. New variables are often included, e.g. debt finance levels, but in general do not demonstrate a consistent significant effect.

By the end of this group, the general case of disclosure in primarily Anglo-American accounting systems was well-studied. Later changes in the overall disclosure environment caused by new regulations or a shift in information demands may lead to new conclusions in later years, but the determinants of disclosure in the early 2000s were well-established. The next development was to stop looking at the general case and turn to specific, non-Anglo-American samples to investigate disclosure in other circumstances.

3.3: Group 2 extension

Late in the second group of papers, researchers begin to take their work in new directions. Most of these move away from the standard in some way, usually either looking into specific types of disclosure such as risk information or looking at countries not examined previously. While the bulk of non-US/UK studies and international comparisons fall into this extension of the group, a few came earlier and are mentioned here for comparative purposes.

The discussion of papers that extend the ideas of the second group into new areas has been split into those that look beyond the primarily USA samples into other countries and those that look at specific forms of disclosure.

3.3.1: Different Countries

Most papers mentioned so far have taken their samples from countries using the shareholder-focused Anglo-American accounting model as opposed to the more bank-oriented Continental model. A number of researchers have examined such countries, however, with mixed results that may reflect the varying national contexts involved. The papers using non-USA/UK samples are summarised in table 3.3 below.

Table 3.3: Non-USA papers				
Paper	**Sample**	**Dependent Variables**	**Significant Variables**	**Theories mentioned**
Biddle and Saudagaran 1989	207 from eight countries	Listing on a given exchange	Comparison of home and second listing location rule strictness	Unique theory; companies list where additional disclosure requirements are low
Meek, Roberts, Gray 1995	226 firms across US, UK, and Europe	Disclosure index focused on strategic information	Size + Mult. listing + Debt finance – Industry effects Country effects	Search for international differences
Raffournier 1995	161 Swiss listed firms	Disclosure index	Size + Foreign operations +	Agency and PCH used in justifying variables
Saudagaran and Biddle 1995	459 multinational firms across eight countries	Listing on a given exchange	Comparison of home and second listing location rule strictness	As 1989 paper by same authors
Zarzeski 1996	256 firms across seven countries worldwide	Disclosure index	Size + Foreign Operations + Debt –	Cultural factors influence disclosure decisions
Adams, Hill, Roberts 1998	600 firms in six European nations	Disclosure index focusing on CSR	Size + Industry effects	Legitimacy supported but incomplete
Depoers 2000	102 French listed firms	Disclosure index	Size + Foreign ops + Big X Auditor + Labour pressure -	Interest in French situation, no named theory
Robb, Zarzeski, Single 2001	192 firms across Canada, USA, and Australia	Disclosure index	Size + Country effects	Cultural commonality and disclosures
Archambault and Archambault 2002	621 across 33 countries	Existing disclosure score system	Mult. listing + Foreign ops + Big X auditor + Cultural factors	Cultural effects on disclosure decisions

Lopes and Rodrigues 2007	55 Portuguese firms	Disclosure index	Size + Mult. List + Big X Auditor + Industry effects	Some indirect Signalling support
Dobler, Lajili, Zeghal 2011	160 firms split between UK, Germany, US, Canada	Disclosure index based on risk sentence count	Size + Country effects	Comparison of countries
Jiang, Habib, Hu 2011	103 New Zealand firms	Bid-ask spread as disclosure measure	Size + Stakeholder –	Ownership concentration effects on disclosure; Agency links

Depoers (2000) ran a typical analysis on a French sample. The results are in keeping with the Anglo-American model papers in many respects. Size is an important and positive determinant of disclosure and a Big 6 auditor has a positive effect. However, perhaps due to the accounting system, debt finance and shareholder numbers have no particular impact on disclosure.

Raffournier (1995) offers a similarly typical test focused on Switzerland. The country is chosen because of an existing gap in knowledge relating to the nation, and the time period studied is just before a tightening of rules. There were few requirements for disclosure before the change and as a result almost all disclosure could be considered voluntary. The observed results are that size and international operations are the only significant variables. Size is then removed from the model. While this removes an important variable, the authors believe it may act as a proxy for other potential determinants of disclosure and could dilute the effects of other variables. However, this alteration does not change anything. While the significance of some other variables rises, they are only significant at the 10% level and not the stricter and more commonly used 5%, making them potentially significant with further changes to the model but still not clearly supported.

Lopes and Rodrigues (2007) study Portuguese practices. The authors are very clear about the fact that their sample is quite removed from the USA context in which the research format was designed and the theories of disclosure were developed, so different results should not be problematic. Despite this, the results are mostly in keeping with other papers, the main point of difference being that a corporate governance variable (measured primarily as independent directors on the board, with another two measures based on independent recommendations for better governance) has been used and found to be insignificant. Note that the research uses a small sample of only 55, however.

As different national contexts were examined, some researchers inevitably tried international studies comparing two or more different countries. An example comes in Adams et al (1998), although the focus here is on the company's social and environmental reports alone. Further, the main determinant of disclosure examined is the home country. This makes it very difficult to compare this paper with others, although a strong and positive size effect is noted.

Meek et al (1995) is a more typical example, building on standard tests with a multiple-country comparative aspect. The sample is split into firms based in the USA, UK, and (continental) Europe, though the entire sample consists of multinationals and renders the categories harder to distinguish as some operate in more than one of these regions. Multinationals are preferred because their voluntary practices have allegedly pre-empted actual regulations in many cases, so this examination is considered indicative of what the future might bring in terms of disclosure regulations. Despite this slightly unusual focus, nothing especially different is found compared to other papers. The company size and (international) listing status remain major determinants of disclosure, although home country is also significant. Industry is additionally found to be somewhat significant, although it depends heavily on the type of information being disclosed. For example, oil companies commonly release more environmental information than other industries.

Dobler et al (2011) perform one of the most wide-ranging international comparisons. While their intent is mostly to compare the countries involved – the UK and Germany as representations of Europe and the US and Canada as representations of North America – the methods are standard and demonstrate the usual results occurring in all four locations, albeit with some national variations. On the whole, US disclosures are the most detailed, Germany is second, and for some specific information the UK gets most detail (specifically environmental information and risk management systems).

Biddle and Saudagaran (1989) (and the same authors' very similar 1995 work) offer a different form of international comparison. Their paper focuses solely on the multinational listing variable as an explanation for disclosure. The usual justification for multiple listing as a determinant of disclosure is essentially companies taking the path of least resistance. For example, a firm is listed in two countries. One requires as part of its listing requirements the disclosure of items A, B, and C, while the other requires B, C, and D. Between the two, the firm has to collect and produce information on four items. It is thought that the cheapest approach is to produce one information package that covers the requirements of both rule sets, leading the firm to effectively disclose voluntarily in each location.

Biddle and Saudagaran (1989) suggest a different relationship, however. Instead of multiple listing generating added disclosure relative to each set of rules, a company will tend to cross-list where the added disclosures will be minimal compared to its home area. To test this idea, the authors examine the cross-listing decisions of a number of companies across eight countries (all in North America and Western Europe). In between these two extremes, the main observation is companies generally selecting to list in locations that are less strict than their home with regard to disclosure regulations. However, note that the US companies should not be considered representative of non-US firms; these companies do not have the option of moving somewhere stricter in this sample. Additionally, little is made of the fact that some companies voluntarily list outside of their home nation for other reasons, e.g. listing in New York after a period of expansion into the US.

A notable oversight here is that the theory being formed is based only on the home nation and not taking all disclosure requirements into account. For example, a Swiss company (historically minimal disclosure) finds a need to list in the USA (maximal disclosure). The theory as written holds that this company would still be unlikely to list anywhere else because of the additional regulation this would impose compared to its native Switzerland. However, at this point the company has to meet the comparatively strict US regulation and may then be willing to report anywhere else as this adds little compared to the company's current regulatory burden.

Related to this, Lang et al (2003) have a very clear finding that US listing leads to analyst following, which is often taken as a sign of disclosure (see Lang and Lundholm 1996, discussed in 5.2.7). It is less clear whether this is due to the American rules on disclosure leading to greater analyst interest or just the fact that the USA is a very large market with a well-developed financial sector combined with a shareholder-centric accounting system. It is possible that any of these factors will lead to a greater analyst following.

3.3.2: Specific Forms of Disclosure

As mentioned above, some papers examine different forms of disclosure. These are summarised in table 3.4 and discussed below.

Table 3.4: Specific Cases

Paper	**Sample**	**Dependent Variables**	**Significant Variables**	**Theories mentioned**
Chen, DeFond, Park 2002	2551 US	Existence of balance sheet forecast	Profit – Ext finance + Volatility +	Disclosure when earnings uncertain
Watson, Shrives, Marston 2002	313 UK firms	Logistic probability of ratio disclosure	Size + Industry effects	Agency and Signalling tested,
Field, Lowry, Shu 2005	166 US lawsuits over information releases	Litigation risks	Profit + Volatility – Industry effects	Disclosure effects on preventing legal actions
Cheng and Courtenay 2006	104 Singapore listed firms	Disclosure index	Governance +	Board composition and disclosure; Agency overlaps this
Ali, Chen, Radhakrishnan 2007	177 US firms, all family-owned	Accruals quality	Profit – Volatility + Family owner	Disclosure by family owned firms; Agency
Brown and Hillegeist 2007	423 US firms	Information asymmetry	Disclosure + Size + Unusual ones	Disclosure as a reduction in asymmetry
Grüning 2007	60 firms split between Germany and Poland	Existing disclosure index	Mult. List + Industry effects Country effects	Unusual methodology, size influences listing
Armitage and Marston 2008	78 UK financial managers	Interview approach; no conventional dependent	Size + Stakeholders +	Interview approach as main method
Chen, Chen, Cheng 2008	1311 US firms, 4415 reports	Probability of various forms of voluntary disclosure	Profit – Ext. Finance + Governance +	Family firm focus; better in some ways, worse in others

Chen et al (2002) look solely at the inclusion of a preliminary balance sheet with interim reports. The paper shows that it is mostly firms with unpredictable earnings that perform this type of disclosure, which implies that to some extent disclosure is used to appear more favourable to the stock markets (in this case, by preventing surprises). However, no further discussion of the result is made and the paper concludes rapidly once the results are obtained.

In a slight variation on the normal research, Robb et al (2001) look only at non-financial information disclosure. The research is otherwise in keeping with the literature discussed above in sections 3.1 and 3.2, using similar explanatory variables to others. The main problem observed here is the unusual choice of significance level. Where most papers (across this topic, wider accounting, and social sciences in general) use the 5% level, the authors in this case accept the weaker 10%. Many of the findings would not be considered significant under the more stringent rule normally used in practice.

The actual purpose of the paper is to examine disclosures by type, counting multiple forms of information separately (e.g. strategy, production information). The paper concludes that the companies most likely to go against their home country's practices and disclose plentiful information are large and geographically diverse, and located in certain industries. Taking the more typical 5% significance level into account, these are still true but the effects are not as strong. Using the 10% rule means these variables are significant across most or all of the classes of dependent variables, while under 5% they are significant across only a few.

Watson et al (2002) investigate the disclosure of accounting ratios. Despite the narrow focus, the results are largely in line with other research into disclosure. Size and industry classification are the most powerful determinants (listing is not tested here), with profit ratios and debt finance having results in some of the years studied but not others.

Walker and Louvari (2003) look at when disclosure of earnings per share numbers are likely to occur, which they find to be under similar circumstances as when disclosure in general is likely, suggesting that EPS disclosures are part of the wider disclosure environment. Note that this paper adds a few unusual variables, such as the firm's overall risk and the use of intangible assets. However, none of the hypotheses surrounding these are supported. The only unusual variable that the results find important is one that measures the overall level of less specific disclosure, which is found to have a positive effect on EPS disclosure. While the results offer some additional details due to the extra information available, the overall outcomes are still in line with past research; the added details add explanatory power but do not suggest the ‘standard’ results are acting as proxies for something not included in the model.

A very different look at disclosure comes from Armitage and Marston (2008). Rather than the usual practice of testing observable company characteristics as explanations for some observable measure of disclosure and forming or testing theories based on this, the researchers here simply ask managers why they disclose. The results are somewhat unexpected. One of the common reasons hypothesised for disclosure is a lower cost of capital. If the market perceives a lack of information on a given company, it will assume there is some potential risk contained in what it does not know and charge accordingly (Lang and Lundholm, 1996). This information risk premium is easily reduced with additional information.

However, the managers asked in this study mostly state that their aim is a reputation for openness. This may have knock-on effects that lead to a lowered cost of capital for the firm, but the managers are not specifically aiming for this effect. The smaller sample necessitated by the interview approach does make this result less generalizable than others, but the added detail from interviews has raised a number of points not seen in other works.

The paper also contains a reversal of one of the common variable justifications. Many have suggested company performance as an explanatory variable on the assumption that managers in well-performing companies will disclose to demonstrate that their leadership was responsible, while poor performance leads to less information so that blame is harder to assign. One of the common themes in the interviews is the complete opposite: companies with poor performance have dissatisfied shareholders who want to see the managers acknowledge the problems and show how they are being solved, while those with good results are not subject to the same scrutiny. One participant calls silence in bad times as an action that would “look stupid”. Bergman and Roychowdhury (2008) find evidence that firms sometimes disclose with the goal of improving the market’s opinions about the company, which is loosely in keeping with Armitage and Marston’s findings.

A few papers have looked at the disclosure practices of family-owned companies, such as Ali et al (2007) and Chen et al (2008). These are usually separated from others because the family ownership means that a small number of closely related shareholders will collectively have a powerful or possibly controlling interest even if no single shareholder has a significant holding. Further, the family is likely to have at least one member acting in a senior managerial position. This family can therefore be considered a single controlling interest as they collectively hold a large number of votes and hold similar motivations, and the existence of such a controlling interest may in turn reduce the overall disclosure. The argument is comparable to the arguments given for heavy use of debt finance reducing disclosure. Since the debt owners are almost exclusively large and powerful entities that can get information by pressuring managers directly and not rely on them revealing information. The family is likely to effectively be an insider if the managerial members share information with relatives, making the entire group well-informed and not in need of disclosure of information. They act as a limit on the demand for information among those with power over the company in this regard. In addition, each paper analyses corporate governance through a measurement of independent directors on the board, generally finding that more such directors increase the company's willingness to disclose, but family firms tend to have fewer of them than other companies.

Overall, however, the results of tests performed in such cases are not much different to more widely-owned companies. While the presence of powerful shareholder blocks does appear to change how much information is revealed, it is still influenced heavily by the size and listing status of the company. Patelli and Prencipe (2007) examine the similar situation of there being a single dominant shareholder, with the result that size is the only typical variable that is significant. The most important one variable is a rare one, the proportion of independent directors on the board. The given hypothesis is that independent directors can pressure the most senior managers to serve all shareholders

3.4: Group 3: The response to Changes

The third group of papers follows major changes to the disclosure environment in the early and mid-2000s. In the USA, Regulation Fair Disclosure (Regulation FD) and the Sarbanes-Oxley Act (SOX) resulted in papers that compare disclosure before and after one or both came into force. In the UK, the major changes examined in literature in the same period were mostly non-mandatory. The ICAEW investigated how risk and risk management reporting was performed and could be improved (ICAEW, 1999). The Operating and Financial Review document at the time encouraged the reporting of major trends and risks, and was a mandatory part of the annual report for a brief period in 2006, but continued to exist as a form of guidance for best practice.

In addition to these rules, various changes in regulation occurred soon after such as the EU adoption of IFRS. The research wave is effectively a broad series of studies on the specific effects on disclosure following various regulatory changes in the early and mid-2000s. These are summarised in table 3.5.

Table 3.5: Regulation research

Paper	Sample	Dependent Variables	Significant Variables	Theories mentioned
Baginski, Hassell, Kimbrough 2002	751 firms in USA and Canada	Existence and time horizon of voluntary forecasts	Size + Profit – Volatility +	Potential litigation discourages disclosures
Linsley and Shrives 2006	79 FTSE 100 firms	Level of risk in firm; risk sentences in report	Size + Overall risk +	Do risky businesses disclose more?
Linsley, Shrives, Crumpton 2006	18 Canadian and UK banks	Level of risk as above	Size +	Do risky businesses disclose more?
Abraham and Cox 2007	71 UK firms	Risk level; count of risk keywords	Size + Mult. List + Governance +	Agency How risks are disclosed
Taylor Tower Neilsen 2010	111 Australian	Risk disclosure index	Size + Stakeholder + Mult. List – Debt – Ext finance + Governance +	Post-IFRS changes to disclosure

Baginski et al (2002) examine changes in disclosure practices following the passing of Regulation Fair Disclosure. Before FD, companies would often pick a small number of preferred intermediaries to receive information in advance of others. However, one of the new regulation's rules forbade this practice, stating that any information released to any party must be available to all parties immediately, although a few exceptions may still have privileged access. The research question is the obvious one for the time, asking how disclosure practices changed following the regulation.

The test involves a sample of firms from the USA and Canada. It is stated that the two countries are very similar and often identical in terms of business practices, but Regulation FD was an American rule only. Effectively, the Canadian firms are serving as a control group here as they are a comparable population not subjected to the change. The results show a clear change over time. While the American companies have continued to disclose, bad news is revealed almost immediately and all disclosures are skewed towards negative information, which is thought to be a measure against litigation as it makes it difficult for an outside party to claim that even potential bad news is being hidden. By contrast, the Canadian sample is more likely to forecast at all times, but especially if the forecast is positive in nature. However, the research is now somewhat dated and further changes may have occurred to either part of the sample. Further, the paper was published in 2002, meaning the research occurred before the Sarbanes-Oxley Act, which further changed US company disclosures.

A number of papers have, rather than investigating on the general concept of disclosure, focused exclusively on risk disclosures. A wave of such research occurred in the mid-2000s in the context of the UK, driven by changes to the regulatory environment at the time which included new disclosure requirements. Linsley and Shrives (2006) offer an early example of such research, although in practice it is not much different to general disclosure papers, having similar variables and similar results. At the same time, these authors participated in another project (Linsley, Shrives and Crumpton, 2006) which focused on risk disclosure among UK banks alone, which again showed little out of the ordinary. Abraham and Cox (2007), when studying UK risk disclosures among companies in general, find the same results as usual, with some extra information provided by the well-justified additional variables included in their model. Taken with the more generalised disclosure research, these papers demonstrate that risk disclosure is likely considered part of the general disclosure package. A company that discloses risk in detail is likely to be one that maintains a policy of thorough disclosure in general terms.

In more recent years, Taylor et al (2010) have performed a related test in the Australian environment to see whether risk disclosure improved over 2002-2006, a period covering the adoption of IFRS in the country. The general observation is an increase over the period, but the list of control variables includes many of the usual disclosure study variables. In a difference from the norm, size is important but multiple listing is not, and actually appears negative. This is explained as a quirk of the sample, which is entirely made up of resource extraction companies. Such firms have to deal with fluctuations in material prices, meaning that in terms of the risks involved foreign exchange is comparatively minor and as such not likely to change much.

3.5: Non-Disclosure

To conclude this section on why companies disclose, we now look at the opposite. Rather than asking why companies disclose, a few papers instead look into why a company might not disclosure, such as Lundholm and van Winkle (2006). The paper begins on a logical argument that, if it holds in practice, makes it clear that disclosure is generally advantageous to almost all companies. The example is a situation where investors have information about the overall quality of the market, but not the individual companies that make it up. Effectively, the investors know how many "good" and "bad" investment options are available but not which companies fall into each category. Every company is then offered an average cost of capital. In this case, companies that are "good" investments could improve their cost of capital by giving more information about operations, making potential investors aware of the lower risks and obtaining a more accurate, lower cost of capital. Should this happen, the pool of firms still being given the average cost of capital is now known to consist only of those which should be given a higher cost of capital and the average rises appropriately, meaning that some firms in this reduced pool could now receive a lower cost by giving more information about their operations. This pattern repeats until nearly all firms are disclosing plentiful information. In the example the paper gives, only one of the five companies used for the illustration does not disclose, but at that point it does not matter because it alone is used to form the average cost of capital and more information would change little.

Even so, there are cases where companies give little information. Lundholm and van Winkle argue that there are three basic reasons for this, all relating to the managers. The first is that they do not know something; information they do not have cannot be disclosed for obvious reasons. The second is that they cannot tell something, usually meaning that they have the information but lack knowledge about if it is important and useful or not. The third and final reason suggested is that the managers simply do not care about some piece of information, which stems in many cases from their incentives to disclose not being sufficient. Additional suggestions have been made in other papers, however. Gigler (1994) points out that firms are wary about giving their competitors too much information, and competitive harm potential is one of the arguments often raised against heavier disclosure regulations. If a given country demands greater information than others, companies listed in the heavier-burden nation suddenly have to reveal information that their competitors based elsewhere need not discuss, giving some informational advantage to any international competitors.

3.6: Problems Identified In Literature

The quantity of papers available on the determinants of disclosure should make for a comprehensive understanding of the topic. The tests have been largely standardised and applied to a variety of contexts ranging from large-sample international comparisons to studies of specific industries, and examined over a period over 40 years long.

While the variables present a clear story of what defines disclosure, the fit values suggest it is not the full picture. At the start of the line of research, Singhvi and Desai (1971) obtained an R^2 value of 0.434. While this demonstrates that the regression is explaining a large amount of variance in the dependent variable, it is a little under half of the total. There is room for further explanation.

However, with studies from various points in time, different samples, new explanatory variables, and even different means of measuring disclosure, Singhvi and Desai's (1971) results are better than most. Raffournier (1995), in a Swiss study, obtains a comparable R^2 of 0.42. More commonly, results are between 0.2 and 0.4, such as Jiang et al (2011, R^2 0.37), Chen et al (2008, R^2 0.3), and Clarkson et al (1999, R^2 0.216). A few papers obtain lower values; Lopes and Rodrigues (2007) find an R^2 of 0.13, suggesting that the Portuguese experience is less predictable than most. Better fits tend to come from very specific cases. Cooke (1989a) has an unusual result, finding an R^2 of 0.6 in one model despite the study being quite standard for the area other than the focus being the Stockholm stock exchange.

Archambault and Archambault (2003) make an interesting contribution to this discussion because multiple models with different variables are tested. One model uses three common variables – size, debt, and a measure of foreign operations – combined with Hofstede's (1984) cultural factors, obtaining an R^2 of approximately 0.3. Another, however, adds a large number of additional cultural variables to the list and obtains a much higher R^2 of 0.54, a clear example of the "normal" variables providing a useful but incomplete picture of disclosure. Depoers (2000) has a broadly similar result. Her work is a mostly standard disclosure study based on a French sample, but the explanatory variables include barriers to industry entry and potential labour pressure. A strong correlation is noted between size and entry barriers, so two regressions are run and one of the two is removed from each model. Without size, but with both of these unusual variables, the R^2 is a high 0.54. Removing entry barriers to include size, making the study comparable to the general body of research, increases R^2 to 0.65.

Higher fits have additionally been observed where the dependent variable is more specific than general disclosure. Abraham and Cox (2007) obtain a higher than normal R^2 of 0.55 overall, but the study is focused on risk information disclosures and contains a number of variables expected to impact on risk disclosure rather than more general information releases. Taylor et al (2010) obtain an $R^2$of 0.575 when examining risk management disclosures, again with a few unusual variables to better explain the specific form of disclosure. Similarly, Dobler et al (2011) obtain R^2 of 0.48 when studying risk disclosures across multiple countries.

Finally, it appears that some countries may be more predictable than others. Zarzeski (1996) performs an international comparison that includes cultural factors and reports both overall and country-specific R^2 values. The fit for the entire sample is 0.48, suggesting that compared to other papers the cultural measures have added a little explanation. Individual countries vary greatly around this, however. The UK appears representative of the sample, having a sub-sample R^2 of 0.47. At the upper extreme, the fit for Norwegian companies is 0.64, one of the highest identified fits. At the low end, the Hong Kong sub-sample has R^2 0.06, indicating that disclosure in this location requires entirely new explanations.

The low fits obtained in many papers may be due to variables that cannot be readily measured. Gibbins et al (1990) examine factors that may encourage or discourage a company from engaging in voluntary disclosure. While some measurable company characteristics such as size and listing status have an effect on disclosure practice, the remainder of the explanation of variance may lie in additional concepts that are not readily measurable, such as the company's history and use of consultants as suggested in this paper.

Kim (1993) has another alternative explanation. Using a modelling approach, the paper suggests that the company will ultimately disclose to the level that its shareholder base desires. This desire depends on the nature of the shareholders. Each individual shareholder can have high or low risk tolerance, and independent of that may have cheap or expensive data collection. Those with low tolerance will want as much information as possible and will therefore press for information releases, especially if they fall into the expensive collection category.

In addition, there are issues beyond low fit in many papers. In general terms, there has been a US focus in the research to date. While other nations and international comparisons have been performed, there have been a disproportionate number of papers using only the USA. This can be explained to some extent by the US accounting system. Where financial reporting is aimed at the banks and similar generally institutional owners of a company, the release of information is not as useful as it would be for small individual shareholders.

In addition, there are questions over the now-standard methodology. The paper commonly cited as the first, Singhvi and Desai (1971), came under immediate criticism from Buzby (1975) for using two correlated variables in a regression. While Buzby (1975) is to some extent criticising what he sees as unexpected results, the core of the argument has merit. It was later backed up by Ahmed and Courtis (1999), who observe signs of a moderation relationship with some of the variables commonly used. The typical testing method of OLS regression should therefore be called into question as it assumes that all explanatory variables are completely independent of one another. While in practice small violations of this rule are not problematic and known larger ones can be worked around, there is little sign of any awareness of this problem in research to date. The extent of any problems caused by this is unknown.

Grüning (2007) recognises the problem and attempts to resolve it. In many regards, the research is typical of the area. A measure of disclosure derived from annual reports is defined and applied to the sample firms. The sample is drawn from Germany and Poland, a little unusual given the US bias in the overall line of research but not a flaw by any measure. However, Grüning questions the use of regression analysis, mentioning co-linearity of variables as an argument for other approaches. In particular, Grüning argues that size and listing status are too strongly correlated to be considered independent.

Between the questions over the appropriateness of regression and the model being something it cannot handle, Grüning instead uses structural equation modelling (SEM). This is effectively a more general case of regression usable where the model does not meet the explanatory variable independence assumption. One of its main flaws is that where regression has the convenient R^2 fit indication statistic, there is no clear equivalent for SEM. Various alternatives have been suggested, and some papers recommend using a range of fit measures for a rounded view of fit (e.g. Marsh et al 1988, Bollen 1990, Shook et al 2004). Grüning does not report on fit to any notable extent, however, leaving the quality of the model in question.

The model tested uses the common variables of size and industry class. Multiple listing is included but in the unusual form of being listed in both countries involved only. In addition, a home country effect is examined by including this as a variable. However, size is not allowed to act on disclosure. Instead, it is considered to be a cause of listing, meaning size may have indirect effects only by acting through another variable. Given past research results, there is little reason why size could not be allowed to act both indirectly and directly on disclosure.

The results of the model support the connections between size and listing and between listing and disclosure. That is, the size of a company influences its listing decisions and this is the major cause of disclosure activity. However, Grüning expresses some misgivings about the results. This is not a modelling problem, but a data collection one; like many, he has measured disclosure by examining annual reports for a number of items that appear in a disclosure index. Grüning is concerned with the subjectivity of this approach, which requires judgement as to whether or not a given item is in the report, and has since moved on to researching ways to solve this problem with automated search procedures, such as his 2011 work on a method known as Artificial Intelligence Measure of Disclosure.

Al-Tuwajiri et al (2004) express similar concerns when studying social and environmental disclosures. The knowledge that some of the variables are endogenous – that is, to some extent caused by others in the model – leads them to conclude that OLS regression is not appropriate. Rather than SEM, they use three-stage least squares regression. This allows for indirect influences, though not models on the same complexity as SEM. The results here support the authors' prior estimate that several key factors in environmental reporting actually influence each other and as such alternative methods are required, but say little about what influences disclosure in any respect.

3.7: Conclusion

In the study of the causes of disclosure, there is a gap in methods. Regression works to some extent, but there is some concern that the variables involved are not independent of each other. This need not be a large problem. Small violations of the independence assumption are not necessarily problematic as this means the results of the model are slightly inaccurate, but likely not enough to make the results unreliable unless some variables have a p-value near to the chosen significance level. In addition, techniques exist such that a link between variables known in advance can be worked around, although this is rarely seen in the research to date. However, this is reliable only if the variable correlations are small and/or known. For many papers, correlations are unknown; the analysis has either not been performed or not been reported.

A minority of papers have discussed potential multicollinearity in the variables. Some, such as Abraham and Cox (2007) and Gelb and Strawser (2001) each find little evidence of it among their samples. In other cases, the possibility of multicollinearity is given as a reason to take or not take a specific action, such as Depoers (2000) excluding one variable from each model because of the potential, although the paper includes no analysis that would confirm the presence or absence of multicollinearity.

In addition to variable correlation, there have been a few rare papers examining the possibility of mediation or moderation effects. Ahmed and Courtis (1999) offer the clearest example of this following a meta-analysis of other research. There is some evidence that variables commonly used to investigate disclosure are acting through each other and having indirect effects on the disclosure variables used as the dependent in the literature.

The variables identified and commonly used are all valid. While not all of them demonstrate consistent effects on disclosure, each has been tested multiple times across a range of studies. Further, and more importantly, every one of them has a good justification for inclusion in a study of disclosure. Without exception, all variables can be argued to have an effect on disclosure. They need to be studied in practice in order to determine the truth, as the justifications are expectations that may not be true. Further, some are sensible arguments on the assumption that a given theory explaining disclosure is valid, but would have no effect if a different explanation of disclosure activity is true.

Grüning (2007) has a good approach with the use of SEM despite the problems in the paper. With care, SEM would be a useful method for solving the problems seen so far. Some means of reporting fit would be necessary to make for useful results.

The extent of indirect effects among variables has barely been examined, although there is a small body of evidence that some exist. Analysis of this would be best performed using multiple models. If one allows variables be connected while another does not, some useful information would be obtained from the difference in fit between the two. Alternatively, methods exist that would allow models to be altered. One model allowing all variables to be connected could be drawn up and connections removed as they are found to be insignificant. A third option exists in that there are ways of identifying links that should be added to a model after it has been analysed, allowing for a basic model that initially allows no correlations to be altered to include those correlations identified as important.

Chapter 4: Theory

This chapter covers the theories of disclosure that are tested later in the paper. Five theories explaining voluntary disclosure by companies have been selected for study. All five are theories initially formed to explain some aspect of accounting or finance other than disclosure, but that have been noted to have potential applications in explaining disclosure.

Other explanations are available, mostly relating to various immediate advantages a company may obtain in capital markets by releasing information. Healy and Palepu (2001) cover many of these in detail, as discussed in section 2.1. These effects are not chosen for study here because they can generally be explained as an effect of one of the theories that is studied. For example, one of the effects of disclosure is a lower cost of capital. Healy and Palepu refer to other research which suggests this happens because of an information risk in the event of non-disclosure; investors with imperfect knowledge have a risk of inaccurate forecasting and require a higher return to compensate. With disclosure, the risk lowers and the required reward falls appropriately. Ultimately, the reduced cost of capital can be explained as an effect of reduced information asymmetry between company managers and investors. Several of the theories discussed in this chapter (Lemons and Signalling in particular) rely on managers reducing information asymmetry and may therefore explain the effects described above. The selected theories are each broad in nature, having been applied widely, and may provide broader explanations of other observations.

Further, a given model result may be consistent with more than one theory rather than highlighting one as a clear best explanation of disclosure. For this reason, the final interpretation of results is not necessarily as simple as finding the one best-fitting theory. To allow for this, the theories are all compared against each other following Morris' (1987) example. The method used has four possible outcomes for comparison. Two theories may be identical, usually alternative wordings of the same observations. Alternatively, one theory may imply the other, suggesting that one is a subset of the predictions of the other. It is possible for two theories to be completely unrelated, in which case one being demonstrated to have some explanatory power does not imply that the other either has or lacks its own power, and both may be valid explanations.

Finally, two theories may compete, in which case they cannot both be explanations of the events as one being true implies the other is false.

The next section explains the method of comparison used later in the chapter. It is explained first because it informs some of the decisions in what is included elsewhere in the chapter, primarily the need for conditions of each theory (see below).

Sections 4.2, 4.3, 4.4, 4.5, and 4.6 each explain one of the five theories in detail, ordered by the date of first publication. Each theory is described as originally written followed by how it has been applied or adapted to explain disclosure behaviour. Next, literature providing evidence either supporting or rejecting the theory as a disclosure explanation is discussed. The following sub-section discusses the expected signs of each theory in terms of the variables used in the study, i.e. the evidence that would support the theory if observed in testing. The final sub-section of each theory section explains the necessary and sufficient conditions of each theory as this is vital to the theory comparison method employed.

Section 4.7 uses the comparison technique to compare all pairs of theories. Two pairs are immediately identified as being identical (at least in practical terms regarding disclosure or due to an inability to model them sufficiently differently with the data used in this thesis). This reduces the number of required comparisons; if theories A and B are identical, then comparison of each to C should have the same results.

4.1: Theory Comparison Method

Morris (1987) provides a means of making detailed comparisons of theories. Any theory can be broken down into a list of the necessary conditions and sufficient conditions required for the theory to hold. These lists can be compared and any overlap between the two offers insight into the relationship between the two. Morris (1987) uses a worked example using Agency and Signalling Theories in a disclosure context, making it a directly useful example in this thesis.

For clarity, the terms ‘necessary condition’ and ‘sufficient condition’ are defined here. A condition X is necessary for some outcome Y if Y cannot occur unless X is true. It is possible for an outcome to have several necessary conditions. For example, a square is defined geometrically as having four sides of equal length and angles of 90^{o} at all four corners. All conditions contained within this statement are each individually necessary for a shape to be a square, but all three are needed to define a square.

A condition X is sufficient for outcome Y if Y being true guarantees X is true. Reversing the example above, a shape being square is sufficient for it to have four sides of equal length. As with necessary conditions, it is possible for several conditions together to be sufficient when individually they are not. Using the same example again, knowing a shape has both four equal sides and four corners of 90^{o} is sufficient to conclude that it is a square, but either condition alone may describe another shape. Note that condition X being necessary for Y means Y is sufficient for X.

Morris’ (1987) method compares the conditions of two theories and allows one of four relationships between them. According to what is termed the axiomatic approach, two theories are identical if their necessary conditions are identical and at least a subset of sufficient conditions is also identical. In this case, the theories are alternative descriptions of the same sequence of events and will make the same predictions; one being true means the other is true, while one being false means the other is also false. Both may therefore be explanations of the underlying observations as they are functionally identical.

A subset match in both sufficient and necessary conditions creates an implication relationship; A implies B if the necessary conditions of B are a subset of those for A and the same is true for sufficient conditions. Theory A being demonstrated to be either true or false means B is the same, but B does not provide information about A in the same way. In this case, both theories may be valid explanations but it is possible that one is valid and the other is not.

Theories are consistent as long as the conditions (both types) do not conflict but otherwise do not have either of the relationships above. Neither theory provides information about the other; theory A being true or false does not have any implications for theory B, and B similarly provides no information on A. In this case, the two theories may both have some explanatory power over the phenomena examined, but will likely explain different aspects of the observations.

Finally, theories compete if there is at least one conflict among the conditions. In this case, theory A being true means theory B must be false, while A being false implies B is true. The two theories are incompatible alternative explanations of the observed phenomena and cannot both be true.

4.2: Lemons Theory

The origin of Lemons Theory is a paper by Akerlof (1970) focusing on the effects of quality and uncertainty within a market. Akerlof uses the example of the market for second-hand cars for clarity, but makes clear that this is an aid to understanding rather than a strictly realistic example. The example simplifies the used car market greatly as it assumes all cars are identical other than being either used or new and good or bad quality (independent of new/used status). No further details such as brand, model, age or service history are included.

The word "lemon", in this context, is a slang term for a defective car. It can refer to a poorly-designed model of which all examples have faults, but in most cases the term is used for faulty individual cars of a normally acceptable type. Two other terms for the faulty examples are "Monday morning cars" and "Friday afternoon cars", each based on the assumption that the car in question was built while the assembly line workers were thinking about the weekend rather than focusing on their work, with the result being a car that has many minor problems that collectively cause it to spend a large proportion of its existence being serviced rather than driven by the owner. These lemons are the bad quality cars mentioned above. In terms of the theory, the important part of a lemon is that the buyer only knows the car is one after the purchase is made. Before purchase, a potential buyer cannot determine whether a car is a lemon or not, as it takes time of owning and using the car for problems to become apparent.

Should the owner of a lemon decide to sell the car, some information asymmetry enters the market. The seller knows that their vehicle is a lemon, but as above a potential buyer lacks this knowledge. The seller has an incentive to set the price equal to that of a good car to obtain the maximum return. Further, the value of a second-hand car cannot be equal to that of a new one of the same type, or the owner of a lemon could sell their faulty car and use the proceeds to immediately purchase an identical new car in the hopes of it not being a lemon this time. The potential buyer, however, is aware that the car in question may be a lemon that the buyer is attempting to replace.

The result of this relatively simple example is that all used cars are sold at one price and, due to the potential that a given car is a lemon, this is lower than the value of a new car. This creates a new problem for the owners of good cars: they cannot get a fair price if they sell. In this simplified example, a good used car is functionally identical to a good new one, but the price for all used cars is lower than that of a new car as it is set between the fair price of a used lemon and that of a new car of (assumed) higher quality. They will not receive a fair price for a perfectly functional car and are less likely to try selling on the open market at all. In this way, the lemons tend to drive the good cars out of the market.

The effect becomes stronger if there are more grades of car than lemons and good ones available. For example, if the set of lemons can be subdivided into those with minor and major defects. The argument above tends to drive the owners of non-lemons out of the market, leaving only the two grades of lemon available. With few good cars available, the price should now be set somewhere between the fair price of a minor lemon and that of a major lemon. Minor lemon owners now have the same problem faced by the owners of good cars in the first example; the market price is below the fair value of their vehicles, which tends to push them out of the market. With an infinite range of quality, the same effect occurs more often; each time the price is set to some weighted average, those with cars worth an above average price have little incentive to sell and leave the market, lowering the average and creating a new set of owners unable to get a fair value.

In practice, car market participants would be interested in additional information beyond whether a given vehicle is new or used. Including such information does not change the logic, however. Rather than all cars receiving a single price, all cars that are similar under all information involved receive one price. For example, all three-year-old red cars of a given make and model and comparable mileage would receive similar pricing. Further, the assumption that a used non-lemon should be identical to a new car is unfair as a used car will have some amount of use and the resulting wear on components, but the existence of lemons means the price will be much lower than is fair under conditions of perfect knowledge.

The important part of the example is the way that information asymmetry favours the seller by obscuring which examples are good purchases and which are bad, and the resulting averaging of prices that may drive the better purchases out of the market. The theory from the paper is that the quality of items in markets deteriorates due to this effect, leaving only the worst options available.

4.2.1: Lemons in disclosure

In capital market terms, Lemons Theory suggests a steady decline in the quality of investments in the market over time. When compared to the used car example, cars are replaced with listed companies, sellers with managers in the companies, and the buyers are potential investors. Managers in a company have the most knowledge of the company's potential for the future and its overall quality as an investment. The potential investors set a price based on their limited knowledge, which tends to be similar to that given to comparable companies and slightly lower than the fair value of a good investment. The sequence of events explained above for used cars then occurs; genuinely good investments are undervalued and, unable to get a fair price (or having the market demand a higher rate of return) than the firm's quality warrants, the company's owners leave the market.

However, if the managers know the company has potential not reflected in its market price then they can communicate this knowledge. Any evidence of the company's quality can be made available to the market. If details on the specifics of a given company are made available in this way then the market price for the specific firm is not subjected to average pricing to the same extent as others and may obtain a value that encourages the firm to stay in the market. Effectively, by disclosing more information, a given company can avoid being considered a potential lemon and suffering the low valuation that would result (Meek et al, 1995). Healy and Palepu (2001) state disclosure to be a potential solution to the lemons problem threatening to harm the entire market.

4.2.2: Research evidence

Lemons Theory is the least researched of those examined in detail in this thesis. The theory is often named as a possible explanation of voluntary disclosure in papers that do not otherwise attempt to test the theory. Core (2001) mentions a "lemons equilibrium", a situation where no disclosure occurs. Akerlof's (1970) paper is named directly in explaining that adverse selection costs mean that a lack of disclosure is costly to companies and may be a cause of demand for disclosure regulations. These two comments represent the total direct mentions of the theory in the paper. Similarly, Beyer et al (2010) name a lemons problem briefly in their introduction and otherwise make little use of the theory. As noted above, Healy and Palepu (2001) note a few possible solutions to the lemons problem including disclosure, but like the others mentioned here make little further use of the theory.

Meek et al (1995) is a rare example of a paper taking the theory beyond a background matter. The paper investigates the determinants of disclosure in multinational companies. The lemons problem is mentioned when discussing profitability as a variable; profitable firms have an incentive to disclose in order to obtain fairer costs of capital. The regression later in the paper finds profitability to be insignificant, implying that Lemons Theory is not supported in this case.

4.2.3: Variables Relevant to Lemons Theory

The basic argument behind Lemons Theory is that firms which are better investment opportunities are likely to disclose more than others. The main sign of support for the theory would therefore be a significant positive effect on disclosure from performance measures. Good performance measures may be caused by competent management (inter alia). High volatility means a company cannot offer consistent returns and would be a less attractive investment, creating an expectation of a negative effect on disclosure from this variable.

Other variables may have less direct effects. A larger company has more capability to diversify into additional lines of business or geographical areas, making it better able to absorb a loss in one area or survive a temporary market downturn. While neither is a clear sign of good management, they each enable the firm to offer consistent returns where a more restricted company cannot, making size a possible determinant of disclosure under Lemons Theory. Similar arguments apply to multiple listing, which is often a sign of international operations.

Debt finance and industry sensitivity are not expected to have effects explainable under Lemons Theory.

4.2.4: Conditions for Lemons Theory

Based on Morris' (1987) analysis of other theories, the following list of sufficient conditions is formed for Lemons Theory:

L1: All market participants are rational wealth maximisers.

L2: The quality of firms competing for equity and debt funds in the capital markets varies.

L3: All firms operate in two periods. Managers make production and investment decisions in the first period which affect the quality of firms in the second period.

L4: The actual quality of firms is observable in period 2 only.

L5: In period 1, information asymmetry exists between the manager of each firm and capital suppliers. The manager's information about the firm is superior to that of capital suppliers.

The only truly necessary condition in this group is L5, information asymmetry. Condition L1 is included because it explains most of the behaviour Akerlof (1970) described as leading to Lemons situations. Capital suppliers aiming to maximise wealth will make investments at the lowest value they consider reasonable, while managers gain from high firm values and are incentivised to claim their company is not a lemon even if it is one. However, it is not necessary on its own; rational wealth maximisation does not imply that the theory holds. Condition L2 allows variation in quality between firms, while L3 and L4 allows variation over time within a single firm while making it unclear which are the best investments until after the investments are made.

4.3: Signalling Theory

Spence (1973) first described Signalling Theory in terms of the job market, which is described in the paper as "investment under uncertainty". Ideally, an employer would pay each employee an amount reflective of the individual's marginal addition to overall productivity. However, this is not knowable before hiring the potential employee. Even after hiring the true productivity may be unclear for a time as the employee takes time to learn the exact details of the position or receive specialist training and will not contribute fully in this period.

An employer unable to know an individual's rate of productivity instead generalises from other characteristics and offers an average wage. A talented employee will be undervalued by this and has an incentive to make their abilities known before the hiring occurs in order to obtain a higher wage from the start of their employment (similar to Lemons, the better employees are offered an unfairly low wage based on averages). Signalling is the act of a job seeker adjusting their observable characteristics.

Signalling is often costly. Education is a clear signal of ability; an employee who has obtained a relevant qualification can be reasonably expected to be more productive than one who has not. However, education is not free. Even where there are no direct tuition costs, the time and effort spent obtaining the qualification carries significant opportunity costs as the potential employee could be earning from a full-time job during the period spent in education. Signalling becomes a case of potential employees attempting to maximise the difference between the wages offered upon entering the workforce and the costs of the signals required for the wage increase.

Spence (1973) makes very clear that the theory requires signals to be costly. If all jobseekers can invest in a given signal to the same extent and obtain the same results then the signal carries little meaning. Continuing the education example, a qualification that is hard to obtain will only be worthwhile to those talented or skilled enough to pass the required assessments. For others, it would be time (and possibly enrolment/tuition costs) spent on failure. A variable level of qualification serves a similar purpose by allowing different outcomes for a fixed cost; a university degree takes the same length of time for most students, but different awards at the end differentiate between skill levels.

In addition, Gibbins et al (1990) note that a free signal allows for false claims since there is no immediate penalty for a false signal. However, Watson et al (2002) argue that even without costs signals should be reliable. A manager caught falsely signalling high quality will lose credibility and further signals will be treated sceptically.

4.3.1: Signalling in disclosure

Signalling theory is directly applicable to corporate disclosure. Spence's (1973) example had an individual signalling to an employer, likely a company. In disclosure, the roles are reversed as a company signals to potential investors who are likely to be individuals.

Signalling as an explanation of disclosure carries a similar starting point to Lemons Theory. Managers in a company seeking investment have the most reliable information on the fair value of the company. Investors lacking this information cannot give perfect valuations to each company and instead value each based to some extent on generalisations. The resulting values will be unfairly low for some of the better investment opportunities on the market. The release of some of the managers' privileged information should enable the market as a whole to give a more accurate, higher valuation to the better investment opportunities.

The requirement for a cost of signalling is still present. The process of disclosure will cover this requirement in most cases. In order to make information available, it needs to be gathered and published. In a company, this means at least one employee must spend time devoted to disclosure that could otherwise be spent being directly productive. Disclosure is therefore only economically useful if it is expected that the market reaction generates enough firm value to outweigh the costs involved.

A second interpretation of the theory in disclosure terms has been used in a few papers (e.g. Healy and Palepu, 2001, and Dobler et al 2011). The company's managers may be signalling not the company's status, but their own competence as managers. This version of the theory has the same overall effects as the original version; managerial competence is observable from the company being profitable.

4.3.2: Research evidence

Signalling is commonly tested as an explanation for voluntary disclosure. Papers testing the theory directly are varied. Many papers examine a specific form of disclosure (e.g. risk information disclosure) and find results that may not generalise to other types of disclosure.

Ahmed and Courtis' (1999) meta-analysis of research to the date of publication does not name the theory directly, but makes reference to a "signalling hypothesis" in which more profitable firms are most likely to reveal additional information. This hypothesis is not widely supported. Profitability measures have positive effects on disclosure in only a few papers, providing weak support for the hypothesis. A suggestion is given that the hypothesis is likely to hold more power than the results suggest, however. Most of the studies in the analysis use only the annual report as a source of information. If companies use any other channels for good news then a possible effect is overlooked in the research used in this paper. Healy and Palepu (2001), the other major overview paper in the field, make use of the management talent signalling interpretation. Analysis is limited to a comment that there is little evidence either for or against the concept in the literature to the time of writing.

Lev and Penman (1990) discuss a signalling interpretation with regard to the disclosure of managerial forecasts. In this case, the observation is that firms either forecast or do not, and this is unaffected by the presence of good or bad news. A company that forecasts will therefore release both news that increases the stock price and news that will reduce it rather than focus only on the good news as Signalling Theory suggests. However, these results may indicate validity to a managerial competence signalling interpretation; while the release of bad news negatively impacts the company's position, the managers have made clear that they were aware of the expected problem in advance.

Muller and Pénin (2006) discuss the idea of knowledge disclosure rather than company information. The paper is focused on which companies take part in innovation networks, loose co-operations between companies, universities, research institutes, and sources of finance to aid in research. The disclosure of knowledge that would otherwise be held within the firm is seen here as a signal of the firm's capabilities as a research partner in order to attract other members to the same innovation network. Effectively, the company's managers are signalling its capabilities to others in order to build a reputation that attracts potential research partners.

Watson et al (2002) examine Signalling as one of two theories that may explain the tendency for a company to disclose accounting ratios. The results offer little support for the theory. The paper examines both Signalling and Legitimacy Theories and most of the results are easily explained in terms of Legitimacy (see below). The lack of a clear financial performance effect on disclosure gives little support to Signalling. The authors suggest this may be due to timing, however. The data for Watson et al's (2002) study is gathered during a time of recession, which is suggested as having a dominant effect that obscures any other performance effects on disclosure.

Hasseldine et al (2005) examine Signalling interpretations as explanations of social and environmental disclosures. The theory is considered useful as a complement to other explanations, but not sufficient alone to explain disclosure. As in other papers, the performance measure used in this study is rarely found to be a significant determinant of disclosure, suggesting that the theory is not a powerful explanation.

While the evidence above suggests that Signalling is generally not a supported theory, the effect of financial performance measures on disclosure cannot be rejected at this point. Each conclusion comes from insignificant effects from financial performance measures. Such measures are a common variable in disclosure research and effects are mixed. While other papers tend not to name the theory directly, the existence of positive effects from performance in some cases offers some (albeit limited) support.

4.3.3: Variables Relevant to Signalling Theory

The signs of Signalling Theory are very similar to those of Lemons. The difference in the two is largely one of intentions. Lemons Theory, as stated above, sees managers of good investments avoiding an assumption of being a bad investment and being valued accordingly. Signalling, by contrast, has managers of good investments highlighting that they are good to attract investment. In either case, the better investment opportunities are advertising themselves as such.

Accordingly, the expected signs of Signalling Theory are the same as those for Lemons. Positive effects from financial performance are the main indicators of the theory, while a negative volatility effect would confirm that the market generally seeks stable returns. Less directly, the ability to diversify from size or multiple listing aids in performance and may have some limited effects.

Debt finance and industry sensitivity are expected to have little to no effects on disclosure under Signalling Theory.

4.3.4: Conditions for Lemons Theory

Morris (1987) sets out the following conditions for Signalling Theory (p50), considering the entire set to be sufficient:

S1: All market participants are rational wealth maximisers.

S2: The quality of firms competing for equity and debt funds in the capital markets varies.

S3: All firms operate in two periods. Managers make production and investment decisions in the first period which affect the quality of firms in the second period.

S4: The actual quality of firms is observable in period 2 only.

S5: In period 1, information asymmetry exists between the manager of each firm and capital suppliers. The manager's information about the firm is superior to that of capital suppliers.

S6: Signalling costs are inversely related to firm quality.

Of these, Morris states that only condition S5 is necessary. Without information asymmetry existing between managers and capital suppliers, there is no need for any signalling. Wealth maximisation (S1) explains the motivation for signalling; those offering a better-than-average investment (be it shares, their skills as an employee, or anything else) would be content with receiving an average value without this, but knowing wealth maximisation occurs is not enough to predict Signalling Theory. Conditions S2, S3 and S4 collectively allow companies to vary in quality over time while making it unclear which firms are the highest quality. Condition S6 is needed for Signalling Theory to function, but like wealth maximisation does not predict it will occur without other conditions also being present.

4.4: Legitimacy Theory

Legitimacy Theory (Shocker and Sethi, 1973) does not explain disclosure as clearly as Lemons or Signalling above. The original idea was explained in a paper on including the views of wider society when developing a corporate strategy.

At the core of the theory is a basic contract that any social institution must follow, businesses included. There are two important terms to this usually unwritten contract: The institution must provide something socially desirable to the broader society (i.e. be relevant), and it must distribute benefits to any groups from which it derives power to maintain its status as a legitimate organisation. Adherence to this contract is vital to the continued existence and growth of the organisation.

In practice, adherence is more complex than the simple contract suggests. Maintaining relevance is complicated by changes in technology and society, either of which can render a once-vital product or service obsolete. Maintaining legitimacy is complicated by the same changes as these can move the sources of power, plus the need to ensure that those benefiting from the business are socially acceptable, as aiding those considered unacceptable is itself considered unacceptable. This final point is further complicated by the existence of subgroups within society which may have very different views on what counts as acceptable behaviour.

The paper goes on to discuss a means by which companies can balance adherence to the contract with continued profitable business. The method involves identifying important groups and their priorities and how a given action may impact on these. At its core, the method involves making decisions based on weighted expected utility functions with the aim of maximising the total utility of all interest groups.

4.4.1: Legitimacy in disclosure

While Lemons and Signalling theories can be linked directly to disclosure, the link is less direct for Legitimacy Theory.

Legitimacy Theory can be used to explain disclosure due to the existence of subgroups in society with specific interests in particular issues. These interests define the group's view on what is acceptable behaviour and in turn what the company must do to retain legitimacy. Over time, the set of groups that are important to a company will change; as the company changes to remain relevant and society itself is changed, so too will the sources of power for the company. The emergence of a new group, decline of an existing one, or change in the relative importance of each to the company will change the underlying utility function that the company needs to maximise. In order to maintain its position, the company needs to keep track of all of the information of this type. One of the tools available is to communicate directly with the interest groups. The release of information about the company is a possible means to this end as it gives information to all possible groups and enables them to communicate any concerns to the company.

The more common approach is to reveal information in order to satisfy society that the firm is not behaving badly. By revealing more about what is occurring within the company, managers can show that the company does not break the implied contract and information can be directed towards showing specific interest groups that the company is listening to their concerns. The company shows that it is behaving acceptably and should not be the focus of protest groups. This process is often referred to as legitimation.

The alternative application is to try to change attitudes. In this approach, information is released that focuses on the positive outcomes of some action rather than showing how it fits with a particular interest group's definition of good behaviour. Rather than avoiding questionable behaviours, the intention is to legitimise them by showing the results in a positive light.

4.4.2: Research evidence

Legitimacy Theory has had some amount of research attention. While there are many papers that mention it as a possible explanation of results (e.g. Haniffa and Cooke, 2005), there is a body of research that attempts to examine the theory in depth.

Patten (1992) offers a good example. Patten studied the environmental information sections of annual reports of various oil companies before and after the Exxon Valdez oil spill in 1989. Exxon, the operator of the ship involved, greatly increased the amount of environmental information after the incident. However, Exxon was only one of a number of companies that collectively owned the consortium running the pipeline the Valdez was using at the time. The other members of the consortium had similar increases in their environmental disclosures in the period, although smaller than Exxon's. In addition, most of the companies examined outside the consortium had increased their environmental disclosures, although again to a lesser extent. In all cases, Patten's interpretation is that the oil industry as a whole received attention over possible environmental problems and individual companies changed their disclosures in response to an expected focus on environmental conduct.

Watson et al (2002) offer a good explanation of the theory's relevance to disclosure (especially within the disclosure of accounting ratios) and make several references to the theory to explain observed results as potentially being companies undertaking a legitimation process. However, the authors advise against the theory as an explanation of disclosure. While considered a useful one for explaining some disclosure patterns, the theoretical underpinnings can be explained using a combination of Signalling and Agency theories. Signalling is the more powerful of the two in this regard as companies are often considered to signal legitimacy.

Adams et al (1998) offer one of the most direct tests of Legitimacy Theory. The research question is whether corporate social reporting practices are a form of legitimation. The final results suggest that there is a strong effect from Legitimacy Theory in determining social reporting practices, but the process is complex and cannot be fully explained by the theory alone.

Lopes and Rodrigues (2007) make minor use of the theory in a study of Portuguese practices. Legitimacy is one of several theories named early in the paper and discussed throughout. However, it is not one of the main points of the paper, and ultimately the evidence is lacking.

Legitimacy Theory is used to justify the inclusion of an industry classification variable due to some industries having more pressures than others. When analysed in various models, this variable is either insignificant or significant only at the 10% level, suggesting it has at most a weak effect within the sample. However, this point is not applicable to many other papers that have made use of an industry classification variable. While few explicitly name Legitimacy Theory as a reason to include the variable, similar reasoning for the variable is used and in many cases a significant effect is observed (e.g. Meek et al 1995, Field et al 2003). There is some support for Legitimacy Theory as an explanation of disclosure from these examples.

4.4.3: Variables Relevant to Legitimacy Theory

Legitimacy Theory as an explanation of disclosure can be summarised as companies faced with any significant controversy tending towards higher disclosure. The expected effects follow from this.

The sensitive industry variable is the main indicator of the theory. The variable is specifically designed to capture sensitivity to public opinions and captures those firms with great environmental impacts through resource extraction and those that directly deal with the public, as this makes them directly subject to public opinion. Accordingly, a positive effect from the industry variable is the most important sign of Legitimacy Theory in the models presented here. A lack of effect would provide evidence against the theory.

Aside from this, other effects are mostly comparable to those expected in the Political Costs Hypothesis below due to the same assumptions of what will generate suspicion among the public. Size is therefore included as major cause of disclosure under this theory. Multiple listing is linked to size and included for control reasons; a larger company has more resources to set up foreign operations, and a multinational has access to more markets and therefore growth potential. In addition, there is a possibility of a multinational company acting in a manner that is appropriate in one culture but criticised in another in which it operates.

Debt finance and volatility are not expected to have any effects under this theory. Volatility may displease investors, but is not a matter of controversy. Similarly, debt finance may lead to a focus on institutional investors, but is generally not a controversial matter.

4.4.4: Conditions for Legitimacy Theory

The collective set of sufficient conditions for the theory is as follows:

L1: Market participants are rational wealth maximisers. This is not necessarily true for other types of stakeholders in a given firm, who are not market participants with regard to the company in question.
L2: Companies must produce something of value to remain relevant and must distribute rewards to those who granted it power or lose support.
L3: All companies operate in two time periods. Managers make investment and production decisions in the first period which affect perceptions of the company's relevance and appropriateness in the second period.

Condition L2 is the only necessary condition of the set. If companies can act without regard to societal views then there is no need for legitimation methods. As in other theories, condition L1's wealth maximisation is needed to explain the company's behaviour. The goal of the company is to generate profit, which in turn is passed on to its investors to some extent, which is not achievable under this theory without support from other parties. While those uninvolved in a given company may also aim to maximise their wealth, the company is not part of this process and may harm their overall utility in other ways.

Variations on condition L3 are common to all theories in the study, being the condition that allows companies to change the relevant characteristic to the theory over time. Its absence would prevent the theory from functioning; a company that takes steps to improve how it is perceived has wasted time and effort if perception cannot be changed. However, where most other theories follow this with another condition about the characteristic only being observable in the second period, a comparable condition is not needed here. The characteristic in question is the public's current perception of the company rather than a more concrete but immediately unknowable characteristic like its quality.

4.5: Agency Theory

Jensen and Meckling's (1976) Agency Theory is an important theory within the discipline of finance. The theory explains the workings of the firm, which at the time they state was often considered a "black box". They see a company as a nexus of contracts between individuals, many of whom have conflicting goals. Where papers of the time were beginning to argue that wealth maximisation alone cannot explain company actions, Agency Theory suggests otherwise. The difference is that where others looked at the company as a unified entity, Agency Theory considers wealth maximisation at an individual level.

The theory relies on an agency relationship. This is a contract in which a principal (the company's owners) contract an agent (the company's managers) to act on their behalf, which requires the agents to have some power to make decisions. If both parties are assumed to be rational utility maximisers, the agent may not always act in the best interests of the principal as the agent's interests may diverge from those of the principal.

The divergence of interests can be limited through various methods, which Jensen and Meckling (1976) split into two broad categories. Monitoring techniques involve observing the agent in order to limit their capacity to behave in a way that does not benefit the principal. The other category, called bonding, are methods in which the agent is authorised to expend resources in ways that guarantee the agent will either not take actions that run counter to the principal's interests or will compensate the principals in such cases. While seemingly counter to the agent's interests, successful bonding methods are carefully designed to increase their wealth. An example is later given in which a manager with some ownership interest is obtaining detailed financial information for decision-making purposes. If the other owners/debtholders decide it would be useful to publish similar information, it is likely that the owner-manager can provide much of the information at a lower cost than any other party. The cheapest course of action would be for the company to have this information checked by an external auditor. For the owner-manager, wealth is maximised by agreeing to pay the auditor's fees as the alternative, having another party gather data from the beginning, would reduce the company's wealth and therefore the fraction that flows to the owner-manager.

Monitoring and bonding methods both carry costs and may still not result in the agent acting perfectly in the principal's interests. There is therefore an indirect cost to the principal in undertaking either method, which is named as the residual loss. Further, the direct costs can eventually reach a point where the cost of bringing the agent's actions into alignment with the principal's interests is higher than the wealth generated for the principal, making the costs inefficient for the principal to undertake and therefore another source of residual loss. Agency costs are overall defined as the three costs combined: monitoring costs plus bonding costs plus residual loss.

Debt finance receives special consideration in the paper. After some discussion of the three agency costs and modelling of where each is optimal, it is stated that few firms are financed almost entirely by debt. This is due to an entrepreneur with low investment relative to the debt total being incentivised to undertake high-risk high-reward strategies since the entrepreneur will receive the benefits of high rewards while personally suffering only a fraction of the costs of unsuccessful projects.

4.5.1: Agency in disclosure

The application of Agency Theory to disclosure relies on the types of cost. While residual losses are unavoidable, disclosure of information can act as either a monitoring or bonding technique.

Disclosure as monitoring is the easier of the two to explain, and it works in three different ways. The first part of this is that information becomes harder to hide within the company. An inefficient use of resources should become more apparent to the owners if more detail of the company's overall actions formally reaches them. Healy and Palepu (2001) suggest that the existence of contracts to reduce agency problems frequently result in the disclosure of information by managers required to demonstrate that they are acting in shareholders' interests.

Related to this, the second aspect is that managers may become more careful about their actions if they are aware they are being observed, stopping some agency costs just by the existence of monitoring methods such as disclosure. Third, published information can reach many. In this sense, by disclosing information the company makes the monitoring more effective by allowing more people to act as monitors.

The bonding aspect of disclosure is less direct. Some of the ideas above are applicable in that providing more information will make it harder for managers to act in their own interests without being observed (Raffournier, 1995). By agreeing to provide more information on their actions, the managers are effectively bonding themselves.

4.5.2: Research Evidence

Agency Theory has received some research attention in the context of disclosure. The use of corporate governance variables as a possible explanation of disclosure is often linked to this theory. Evidence from research has been mixed.

Watson et al (2002) find evidence broadly consistent with Agency Theory explaining the disclosure of accounting ratios. Even within the paper, there is some doubt. One of the expectations is that the more regulated industries in the sample would tend to disclose more as a means to reduce agency costs, but this particular hypothesis was not supported. An earlier paper, Meek et al (1995), demonstrates evidence that contradicts the theory. A measure of leverage is included and found to be significant, but with the wrong sign. Where the theory predicts highly leveraged firms disclosing more to limit possible agency problems, the finding in the paper is more disclosure from less leveraged companies. By contrast, Ahmed and Courtis' (1999) meta-analysis finds a broad association between high debt finance and disclosure, supporting the theory. Similarly, Abraham and Cox (2007) find evidence supporting Agency Theory. The difference may be explained by the more specific nature of this paper, however. Rather than a general measure of disclosure, the paper examines risk information disclosure only.

Outside of the Anglo-American accounting context, evidence tends not to support the theory. Depoers (2000) makes frequent reference to the theory but rejects it as an explanation of disclosure in France. Similarly, Lopes and Rodrigues (2007) use the theory but, finding few of the Agency-derived variables to be significant, cannot support the theory as an explanation of disclosure. However, Patelli and Prencipe (2007) find evidence supporting the theory in an Italian context, albeit limited only to cases where there is a single dominant shareholder.

4.5.3: Variables Relevant to Agency Theory

To simplify, companies can reduce agency costs by disclosure. This suggests that the signs supporting the theory would be firms likely to experience agency problems tending to disclose more.

Debt finance is the primary indication of agency problems. The existence of both equity and debt ownership means there are two classes of owner with different priorities, meaning a manager will inevitably generate some agency costs by serving the needs of one but not the other. Jensen and Meckling (1976) include a lengthy and complex discussion of how debt may lead to agency costs; while it assumes managers to be dishonest, it gives an example of how debt can add to agency problems in a firm. Despite this, a clear direction of effect is not apparent. The theory applied to disclosure suggests that debt finance creates agency problems and should lead to disclosure. However, lenders tend to be large institutions capable of acquiring information privately. The result is a reduced need for disclosure as a portion of the company's owners has no need for this type of information release. Debt is expected to affect disclosure, but a strong effect of either positive or negative nature can be argued to be consistent with Agency Theory.

The size of the firm is expected to have a positive effect on disclosure under this theory. A larger company offers more opportunities for a manager to act in their own interests. The sheer volume of information the company generates means more information is summarised rather than discussed in detail, potentially glossing over a manager's self-interested actions. Similarly, multiple listing generally means international operations, which means less oversight from the headquarters and has the potential to create language barriers that make information harder to interpret. Again, a positive effect on disclosure is expected as more information is produced to limit these effects.

Clear effects from other variables are not expected. Neither volatility nor sensitive industry membership have large agency cost implications. It is not clear whether Agency Theory predicts a positive or negative effect on disclosure from financial performance. No effects from any of these variables are included in Agency Theory models in this study.

4.5.4: Conditions for Agency Theory

Morris (1987) lists the following as the complete set of sufficient conditions for Agency Theory (p50):

A1: All market participants are rational wealth maximisers.

A2: All firms operate in two periods. Managers make production and investment decisions in the first period which affect firms' expected values and variances in the second period.

A3: Firms have external equity and debt financing.

A4: There is separation of equity and debt finance suppliers and managerial control in the firm.

A5: Each manager owns a fraction of the firm's outstanding equity.

A6: Each manager is remunerated by salary, perquisite consumption, and returns on their equity in the firm.

A7: Monitoring and bonding are available at a cost proportional to the value of the firm; and they reduce activities dysfunctional to capital suppliers at a reducing rate.

Conditions A1 and A4 are stated to be collectively necessary. Taken together, they imply that managers and capital suppliers are separate groups but both try to maximise their own wealth. This allows the existence of the agency costs on which the theory relies. A6 does not imply Agency without A1 and A4, but its direct mention of perquisites explains some of the sources of agency costs. A5, managers owning part of the company, is a common means of performance-related pay used to align managers and owners. A3, a mix of debt and equity capital, reflects the importance Jensen and Meckling (1976) placed on companies mixing the two sources. As in other theories, condition A2 allows changes over time. Finally, condition A7 explains how agency costs can be directly reduced, but as with others its presence does not predict the theory without further conditions.

4.6: Political Costs Hypothesis

The last theory tested in this thesis is part of Watts and Zimmerman's (1978) Positive Accounting Theory (PAT). This, the authors' first paper on PAT, argued for the idea of approaching accounting research from a positivist viewpoint. The focus of the theory is what position a given company will take when lobbying with regard to proposed new rules and regulations, which is argued to be the result of the managers' attitudes.

Political costs are loosely defined as wealth transfers caused as a result of politics. Various examples are given, ranging from the nationalisation of a company to the regulation of an industry. One of the major causes of political costs is explained to be the reported profits of a company, caused by an association the public makes between larger companies and monopolistic behaviour, in turn leading to increased demands for some manner of regulation on the company in question.

The same authors provided a review of Positive Accounting after some time had passed and others had worked with the ideas (Watts and Zimmerman, 1990). Three commonly used hypotheses are detailed in this paper, one of which is named as the Political Costs Hypothesis. The hypothesis holds that larger companies will generally lobby in favour of proposals that would reduce their reported profits. While counterintuitive given the normal assumption that high profits are desirable to both current and potential investors, the reduced profit lowers the amount of attention the company receives from the public and in turn the amount of political pressure directed at it.

Note that the hypothesis above is based on reported profit causing political costs, where this is a measure of size rather than financial performance. Further, a more general version of the hypothesis is that various characteristics make a company more visible to the general public, which increases the chance of the firm suffering political costs. This visibility is referred to as "negative attention" below.

4.6.1: Political Costs in disclosure

In terms of disclosure behaviour, the Political Costs Hypothesis may act in two different ways.

The more direct of the two ways is that disclosure can add more detail to the information available and reduce the negative attention received. Managers in a large company may be able to disclose some additional information that makes the position appear less powerful, such as reporting difficulties faced by the company in the period. Alternatively, engaging in and reporting on programmes that benefit people outside of the company (e.g. environmental initiatives) may reduce negative attention by demonstrating that the company is not solely focused on its position in the market.

The alternative is for managers to accept that negative attention will occur and attempt to limit the resulting regulation. In this approach, the managers observe the calls for regulation that will affect the company and produce relevant information before anything is made mandatory. For example, if the political process seems to be moving towards requiring environmental information, managers may reveal some of this before any rules come into force. The intention is to satisfy some of the demand for information of the type, reducing the political pressure and the likelihood of onerous regulation being placed on the company.

4.6.2: Research Evidence

Despite its status as a hypothesis rather than a theory, political costs have received much research attention in disclosure. Overall, the evidence supports the idea.

Healy and Palepu (2001) discuss the theory in various parts of their literature review. While no conclusion is drawn, the evidence presented generally does not support the theory, finding little to no effect on various measures when political concerns are involved in a study. One part of the paper suggests that size is not a suitable proxy for political costs as it captures many other factors, making any effects observed difficult to definitively explain as the result of political cost concerns. Raffournier (1995), however, states the exact opposite; while size may be correlated with other variables, its effects on disclosure should be attributed to political costs. With this, Raffournier's paper supports the Political Costs Hypothesis in disclosure as size is very powerful and significant. Ahmed and Courtis (1999) similarly use size as evidence for a political costs interpretation of disclosure.

Whether size is an appropriate measure of political costs or not is vital to the theory; if size does indicate political costs, then many other papers support the PCH as an explanation of disclosure. If not, then many papers have shown that size is significant and provided little or no information on the PCH.

Meek et al (1995) do not state a conclusion regarding the theory, but many of the variables they argue may signify Political Cost effects are significant in the final analysis, lending some support to the idea.

Watson et al (2002) offer some additional criticisms of the PCH. In addition to questions over whether size is an appropriate measure, one footnote suggests that the exact opposite of the theory may be true. As written, the theory holds that companies with the potential for political costs should disclose. This paper instead suggests that a company with likely political costs would be better off remaining silent in order to avoid any attention that may lead to costs.

Finally, Milne (2002) argues against the entire concept. While much evidence exists and is consistent with the PCH, Milne questions it. Many papers have used such a broad definition of the term "political costs" that alternative explanations may be valid with the same results. Without a firmer definition of political costs, there is little to support the theory directly.

4.6.3: Variables Relevant to the Political Costs Hypothesis

The PCH is similar in many ways to Legitimacy Theory in many ways. While it is based on negative attention rather than controversy directly, the two concepts overlap heavily.

Watts and Zimmerman (1978) explained when political costs would be most likely to emerge as they introduced the idea and later (1990) clarified it to explain the Hypothesis as larger firms being more likely to experience political costs, making this an important variable for the PCH. As with Legitimacy, multiple listing is included although mostly as a control variable due to its links to size. Further, the Hypothesis is based on possible calls for regulation due to the company attracting negative attention, making the sensitive industry variable a likely cause of disclosure as it covers some of the companies often named as environmental problems.

As in Legitimacy Theory, volatility and debt finance are not significant causes of controversy and not included in the theory.

4.6.3: Conditions for the Political Costs Hypothesis

The collective sufficient conditions for the PCH are as follows:

P1: Market participants are rational wealth maximisers. This is not necessarily true for other types of stakeholders in a given firm, who are not market participants with regard to the company in question.
P2: Companies with a negative public perception will experience various negative outcomes.
P3: All firms operate in two periods. Managers make production and investment decisions in the first period which affect the company's public perception in the second.

P2 is the only necessary condition. Without the concept of negative attention, the hypothesis collapses as there is no need to reduce attention to avoid the undesirable regulation that would result. Market participant wealth maximisation (P1) in included because there is an implicit assumption in the PCH that regulation limits the company's ability to profit, but as in other theories it does not predict the PCH alone.

Condition P3 enables managers to change the perception of their company over time, again important to the PCH but not enough to predict it. As in Legitimacy Theory, the PCH concerns the immediate perception of the company rather than a reliably measurable but currently unknown characteristic and no condition about observing it only in the second period is included. Condition P1, again similar to Legitimacy, means a company's managers and owners each aim to maximise its profit. Those who are not involved in a company through the market (i.e. most stakeholders) do not gain wealth from the company's actions.

4.7: Comparison of theories

The list of variables relevant to the PCH in section 4.6.3 is nearly identical to those in section 4.4.3, the variables relevant to Legitimacy Theory. The only differences in the two are some of the justifications for a given variable; the actual list of variables and the effects expected of them is identical. There is clearly some overlap between Legitimacy Theory and the Political Costs Hypothesis that makes the two impossible to distinguish from the list of variables examined here. However, much of the overlap in relevant variables results from the two explanations having similar underlying assumptions. Each is based on companies attracting unwelcome attention. A difference occurs only at this point; the PCH assumes regulation will be demanded and the company's managers will take actions to limit its effects, while Legitimacy Theory has managers communicating with stakeholders over their concerns about the company's problematic aspects.

The similarity is enough that the two may be identical for practical purposes. While differences exist in the two as explanations, they describe largely the same series of events in disclosure terms, differing mainly in the intentions of the company managers. Lemons and Signalling theories are in a similar position, each explaining why firms with better results and/or managers may publicise more information.

In addition to the possibility that some theories are the same in disclosure terms, the research performed aims to determine which theory or theories best explain observed disclosure activity. It is possible that several have explanatory power and as such a formal means of comparing the theories before analysis is needed to ensure that the results are reasonable, such as not supporting two contradictory theories.

Table 4.1 below repeats the conditions for each theory for reference purposes.

Table 4.1: Conditions by theory
Lemons Theory conditions L1: All market participants are rational wealth maximisers. L2: The quality of firms competing for equity and debt funds in the capital markets varies. L3: All firms operate in two periods. Managers make production and investment decisions in the first period which affect the quality of firms in the second period. L4: The actual quality of firms is observable in period 2 only. L5: In period 1, information asymmetry exists between the manager of each firm and capital suppliers. The manager's information about the firm is superior to that of capital suppliers.
Signalling Theory conditions S1: All market participants are rational wealth maximisers. S2: The quality of firms competing for equity and debt funds in the capital markets varies. S3: All firms operate in two periods. Managers make production and investment decisions in the first period which affect the quality of firms in the second period. S4: The actual quality of firms is observable in period 2 only. S5: In period 1, information asymmetry exists between the manager of each firm and capital suppliers. The manager's information about the firm is superior to that of capital suppliers. S6: Signalling costs are inversely related to firm quality.
Agency Theory conditions A1: All market participants are rational wealth maximisers. A2: All firms operate in two periods. Managers make production and investment decisions in the first period which affect firms' expected values and variances in the second period. A3: Firms have external equity and debt financing. A4: There is separation of equity and debt finance suppliers and managerial control in the firm. A5: Each manager owns a fraction of the firm's outstanding equity. A6: Each manager is remunerated by salary, perquisite consumption, and returns on their equity in the firm.

A7: Monitoring and bonding are available at a cost proportional to the value of the firm; and they reduce activities dysfunctional to capital suppliers at a reducing rate.
Legitimacy Theory conditions L1: Market participants are rational wealth maximisers. This is not necessarily true for other types of stakeholders in a given firm, who are not market participants with regard to the company in question. L2: Companies must produce something of value to remain relevant and must distribute rewards to those who granted it power or lose support. L3: All companies operate in two time periods. Managers make investment and production decisions in the first period which affect the public's perception of the company in the second period.
Political Costs Hypothesis conditions P1: Market participants are rational wealth maximisers. This is not necessarily true for other types of stakeholders in a given firm, who are not market participants with regard to the company in question. P2: Companies with a negative public perception will experience various negative outcomes. P3: All firms operate in two periods. Managers make production and investment decisions in the first period which affect the company's public perception in the second period.

4.7.1: Comparison 1: Lemons and Signalling

The descriptions above for Lemons Theory and Signalling Theory suggest some similarity in the two. A comparison of conditions suggests a stronger connection.

The necessary conditions are L5 and S5 in table 4.1, which in each case is the existence of information asymmetry in which company managers have superior information to potential investors. Aside from the final condition S6, the two sets are identical. Under Morris' (1987) technique, this makes the theories identical.

Each covers a situation where information asymmetry exists regarding the quality of an investment and the potential investors set all prices on the assumption that any opportunity may be a poor one. The primary difference in them is what is suggested to occur after this point. Lemons, as written, has the better investments leave the market and reduce the overall quality. Signalling, by contrast, has the better investments find ways to make their superiority clear to the market. In effect, Lemons describes what would happen if signals could not be sent. Unable to distinguish themselves from inferior investments, the better options cannot obtain a fair price and leave the market.

In practical terms, sections 4.2.4 and 4.3.4 detail the observable effects of these two theories among the variables tested here. Each relies on companies highlighting good performance and, as such, the two would be impossible to differentiate in this particular study. They are treated as the same theory for the remainder of the thesis. Future references will name only Signalling as it is more often mentioned in other disclosure literature.

4.7.2: Comparison 2: Legitimacy and Political Costs

The descriptions above highlight similarities in Legitimacy Theory and the Political Costs Hypothesis. Both relate to the perceptions of companies leading to negative outcomes and the managers attempting to reduce the problems by actions that affect the perception of the firm.

There are clear similarities in the sets of conditions given in table 4.1. The second condition, the necessary one in each, is different. However, while the wording differs, the two conditions describe very similar situations. In terms of explaining disclosure, the two theories are almost identical. They differ in broader terms, but in disclosure terms they overlap.

For the purposes of this study, the two are considered identical. In each, companies likely to come under question are expected to disclose more information. While it is questionable whether the two theories are the same as explanations of disclosure, they are impossible to distinguish using the variables employed here. The two are hereafter referred to collectively as Legitimacy, using the name of the Theory rather than the Hypothesis. Either may be the underlying explanation of any evidence that supports the two, however.

4.7.3: Comparison 3: Signalling and Agency

The comparison of Signalling and Agency Theories is performed in Morris (1987) as an example of the method.

Morris argues that condition S5 is necessary for Signalling, while A1 and A4 are necessary in Agency. That is, in order, information asymmetry, wealth maximisation, and separation of ownership and control. Information asymmetry and separation of ownership and control complement each other since the controlling managers will tend to have more information about their company than any other party. However, Signalling Theory does not have wealth maximisation as a necessary condition, indicating that the two theories are not identical and do not share an implicative relationship. Further, at no point do the conditions conflict. They are therefore concluded to be consistent theories in that neither provides information about the other and each may be true or false independently. Both may therefore have some explanatory power over disclosure in practice.

4.7.4: Comparison 4: Signalling and Legitimacy

Earlier discussion suggests that Signalling and Legitimacy have little in common. Nonetheless, a formal comparison is carried out.

There is no overlap in the necessary conditions of these two theories. The necessary conditions are S3 and L2, respectively the existence of information asymmetry and companies potentially losing support if acting inappropriately. Neither implies the other, and neither conflicts with the other. In the sufficient conditions, only wealth maximisation is common and there is no evidence of conflict among the rest. These theories are therefore consistent; Signalling and Legitimacy Theories may both have some explanatory power over disclosure independently of each other.

4.7.5: Comparison 5: Legitimacy and Agency

As in section 4.1 above, there is little in earlier discussion to suggest Legitimacy and Agency theories share either conditions or predictions.

There is little evidence these theories interact at all. The necessary conditions are L2, A1, and A4; companies must act appropriately or lose support, market participants maximise their wealth, and the separation of ownership and control. There is some overlap in these conditions. Separate ownership and control could potentially lead to the firm not rewarding owners in the event of serious agency problems. However, wealth maximisation is only necessary in Agency Theory and not Legitimacy, where it is a (jointly) sufficient but not necessary condition, so the two are neither identical nor implicative. Examination of the remaining conditions shows no overlap at all, suggesting that these two are consistent theories that may both explain disclosure independently.

4.8: Conclusion

Five theories of disclosure have been selected for study. These are Lemons Theory (Akerlof, 1970), Signalling Theory (Spence, 1973), Legitimacy Theory (Shocker and Sethi, 1973), Agency Theory (Jensen and Meckling, 1976) and the Political Costs Hypothesis (Watts and Zimmerman, 1978 and 1990). Each has been explained both in general and as a theory of disclosure.

In comparing theories, Lemons and Signalling are identified as being identical. While there are differences between the two, as explanations of disclosure they are identical. In Lemons Theory, companies are priced as if there is some risk every company is an unknown lemon. Managers can counteract the potential lemon assumption by providing more information about the company. This is almost exactly what Signalling Theory describes; better investment opportunities make their superiority clear in order to obtain the higher valuations that they deserve. The two are collectively referred to as Signalling Theory in the remainder of the thesis.

In addition, Legitimacy Theory and the PCH are considered identical. This is less clear than the above example and is more a practical than a theoretical effect. Both theories have managers in companies that attract the wrong kind of attention provide more information in order to reduce the suspicion surrounding the company. The signs of each theory are therefore similar enough that, while the theories are not identical, they would be impossible to distinguish in the study performed. The two are collectively referred to as Legitimacy Theory in the remainder of the thesis, but the PCH explanation is an equally valid one for any supporting evidence despite this name.

Finally, Signalling, Legitimacy and Agency are all compared against each other. Morris (1987) performed the comparison of Signalling and Agency and found no evidence that the two were at all linked. The same is found for the other comparisons performed. These three theories are all possible explanations of disclosure independently of each other and some combination of them may explain disclosure behaviour among company managers. The next phase is testing these ideas, which requires data.

Chapter 5: Data and methods

This chapter covers the origins and usage of the sample used in the study and the variables used in later modelling. The chapter explains where the raw data originated and explains which companies make up the sample. The basis of the sample is a pre-made list of companies, so part of this section will explain what this list contains and explains the reasoning for removing some entries before analysis.

The data for this study was mostly gathered within DataStream, primarily using the program for access to the Worldscope database. One of the pre-existing lists of companies in DataStream was identified as a useful starting point, consisting of 1660 FTSE-listed companies at 31/7/2013. However, as in many papers using a broad list, not all were suitable for entry into the sample. The list's constituents are not removed immediately should they de-list for any reason, commonly one of choosing to become private entities, cessation of business activity due to bankruptcy, or merging with another company. Instead, they remain part of the list for some time after their de-listing from the FTSE with a note about why there is no additional data available. The lack of further data makes these 'dead' firms (to use the program's term) unusable.

In addition, some entries were missing important data. The majority of these are investment funds run by companies rather than shares in the companies directly. This would not be a problem in itself, but they generally lack at least one piece of data and are unsuitable for comparison with others. This is likely a result of their nature as investment funds as some of the information is not necessary or useful to these different types of investment. In addition, there are some cases where a company lacks some piece of information.

Between dead firms and missing data, the final usable sample contains 1436 companies. No further restrictions are placed on the sample (e.g. industry sector), leaving this as the final sample size.

The remainder of the chapter is organised as follows. Section 5.1 discusses the the choice of sample, the possibility of outliers, and the transformations applied to the data set before use in some later models. Section 5.2 details the variables used in the study, the individual characteristics of the companies included in the sample that are used in the models performed in the next two chapters. Section 5.3 investigates the variables before modelling is performed, examining the descriptive variables and performing some correlation analysis to inform later modelling.

Finally, section 5.4 details the modelling methods used in the next chapter. The section primarily focuses on the structural equation modelling aspect. While regression methods are used and are important to the study, they are also far more commonly used in disclosure literature and familiar to researchers in the area.

5.1: General Discussion of Data

This section provides a discussion of the reasoning behind the choice of sample.

A large data set is useful in any quantitative test. A sample should be representative of the wider population in terms of the various characteristics the population may have. The larger a sample becomes, the more likely this is to be true. A small sample is more likely to not include a member of the population that represents a particular characteristic. For example, if only 5% of companies have no debt in their capital structure, a small sample may fail to include a single company of this type. Alternatively, a small sample may include proportionately too many of these companies and create the impression that a larger percentage act this way. The larger the sample becomes, the more likely it is to accurately reflect the population's characteristics.

However, feasibility of using larger data sets can become an issue. The larger the sample becomes, the more observations there are and the more unwieldy the data set will be. The same occurs when more variables are used as this increases the data required of each case studied. Whether through more cases or more variables, a large data set means more time is required to both gather the data and manipulate it for study. The data collection technique is especially influential in this regard as slower approaches will be particularly hard to use with a larger sample, but may provide better information on a single case.

The researcher needs to therefore balance the two concerns of practical data sets and representing the population accurately. In this thesis, there is an opportunity to study a population instead of a sample. The data is mostly obtained through Datastream (DS), an online database of company information. This allows large quantities of information to be obtained very quickly, being limited more by internet connection speeds than the need to gather data directly from the source companies. It is therefore possible to gather information on all companies that DS covers very quickly, making the entire population of UK listed companies a feasible data set to examine.

The purpose of the research is primarily to investigate whether SEM is a possible means to handle the rarely-discussed endogeneity present in the data used in disclosure studies. Many potential samples are available for this idea. The sample selected consists of all

UK listed companies as of July 2013. There are four characteristics contained in this description, and the reasons for each are discussed below.

Listed companies are chosen due to a practical data availability matter. Listed companies are required to produce and make easily available detailed financial information annually along with smaller interim releases and occasionally immediate releases of important new information. Acquiring data on listed companies is accordingly straightforward.

Unlisted companies, by contrast, need only to provide information to existing shareholders, although they may provide information on request. This in itself creates a possible source of problems as the two types are subject to very different requirements. Further, if unlisted companies are used, there is a clear potential response bias problem in that not all companies will respond to requests. If there is any consistency in those that provide information then there is a bias in the sample. It may be the case that this bias actually provides information about one or more of the theories of disclosure, e.g. unlisted companies with high financial performance responding to requests for information would be evidence consistent with Signalling Theory. However, without information on the non-responsive companies, it would be difficult if not impossible to determine whether the responding companies are any different to the rest. In addition to possible bias, the process of gathering information on unlisted companies is comparatively slow as unlisted companies do not appear in the Datastream database in any significant number. Between possible response bias and slower information gathering combined with the intention to use a large data set, unlisted companies are excluded. This is consistent with comparable studies, which rarely use unlisted companies in their samples.

Disclosure studies have historically tended to use samples from shareholder-focused Anglo-American accounting system countries. The reason for this is simple: private shareholders should be the primary beneficiary of disclosure. A larger institutional shareholder is likely to have the power to pressure managers into releasing information privately, or otherwise have the power and ability to learn fine details about the company. A smaller shareholder, by contrast, is not able to pressure the company in the same way and may miss out on details not provided in some kind of disclosure. It is likely that shareholder-focused annual reports result in greater overall disclosure than those in countries where annual reports are intended for lenders. As the intention behind this thesis is to create a thesis comparable to past disclosure studies, an Anglo-American system country is the clear choice. While this narrows down the possible countries to use, there are still several options. A UK sample is chosen primarily due to researcher familiarity with the sample.

Early in the research, exploration of the information contained within DS revealed that the program contains a sample data set containing all UK-listed companies. This provided a sample that was large but not unmanageable. Further, as discussed above, there are advantages in being able to use the entire population instead of a sample drawn from it, making this a useful starting point for forming the final sample. The DS sample contains a number of entities that do not provide all of the information required. In addition, this sample is updated as new companies are listed and existing ones de-list, but there is a delay. The effect of this is that there were a number of companies described as 'dead' in the list that needed to be removed. This term refers to any company for which information is no longer available. Common reasons for this lack of current information include cessation of business activity, de-listing, or a change of name (in this case, the old name entry will often refer users to the new one). Mergers and acquisitions may create one or more of these situations among the companies involved and are another common cause of 'dead' companies in the DS sample. The exact effects depend on the structure of the company after the merger or acquisition is complete. Commonly, one of the companies involved increases in size as the final business entity keeps the name of one and combines the assets of both, while the other is considered 'dead' as its name is either no longer in use or does not refer to an independent listed company.

The final aspect of the sample is the timing. Data was gathered in mid-2013 and the decision was made to use up-to-date information at the time. In Datastream, the request was made for the most recent information at the date of 31/7/2013, resulting in information dated from the most recent report before that date, i.e. anywhere from August 2012 to July 2013 depending on the financial year end. This date provides data that is recent in comparison to most disclosure studies and is well after the stock market crash of 2008. The crash represents a potentially interesting period in which to study disclosure behaviour to determine whether companies behaved differently due to the sudden downturn. However, the intention in this thesis is to create a study comparable to the main body of disclosure studies, the majority of which were tested on recent data at the time of publication.

There is an additional practical reason to use data from this time period. The DS sample is updated regularly and its constituents are always currently UK-listed companies. The actual date at which the list was used determines its members, not the time for which data is requested. For clarity, using this list at July 2013 will return data on the companies listed at that point regardless of what historical data is requested. Using this list to investigate 1993 would cause problems due to many of the constituents being unlisted or yet to form in 1993, while many 1993-listed companies would not be included as they now count as 'dead' companies in Datastream. This problem could be solved without much difficulty by making a custom list containing all listed companies that were active at the time for which data is required, however. Using the DS sample is faster and easier without compromising on data other than requiring up-to-date information at the time of data gathering.

'Dead' companies would create a potential survivorship bias in the DS sample in a study over time, but data in this thesis is gathered for a single point in time only and should contain a mix of long-established and stable companies, failing companies, and newly-formed or newly-listed entities. The sample was investigated again at a later date due to some companies returning unlikely values for some of the collected data points. A few instances of negative shareholders' equity are noted (as discussed in section 5.3.7) and there are many cases where large reported losses lead to negative performance ratio values. When the same DS list was retrieved at a later date, many of the companies showing these traits were 'dead' companies.

5.2: Variables

This section discusses the many variables used within the study. All variables are classed into one of seven groups. The reason for this is explained in more detail in section 5.4 below. To summarise, SEM can make use of latent variables, a type of variable that is not observed directly but instead indicated by several others that are observed. Variables are classed here under headings based on which latent variable they indicate, meaning that all variables under one heading are alternative means of measuring the same concept.

5.2.1: Size

A company's size is thought to lead to greater disclosure. This may occur as a larger company will generally have more to report, leading to an economy of scale effect (Singhvi and Desai 1971, Raffournier 1995). As the quantity of data about the firm rises with size, senior managers are likely to receive summaries of important information. The kind of data often disclosed is therefore already collected and collated within the firm's ordinary operations and the only additional costs are in adapting it for external use and publication. Further, there is some suggestion that managers may avoid disclosure due to proprietary information that may threaten the company's competitive position. Singhvi and Desai (1971) argue that this is less common with a larger company as its overall position is harder to threaten.

The vast majority of literature examined includes a size variable. Almost all of those that use the variable find it has a significant and positive effect on disclosure (e.g. Buzby 1975, Adams et al 1998, Linsley et al 2006, Dobler et al 2011), and no negative results have been identified. In general, insignificant results occur when the sample is restricted to specific types of company. For example, Ali et al (2007) and Chen et al (2008) both find no size effect when examining family-owned companies. Malone et al (1993) find no effect when examining only oil and gas companies, although Patten (1992) does observe a size effect in a similar sample. Gelb and Strawser (2001) provide an anomalous result, with size being insignificant despite a seemingly typical sample. The only limits placed on the sample were that firms had to have made available social and environmental information as early as 1989, possibly meaning that the sample is mostly larger companies that had experimented with such reporting early. In a case highly relevant to this study, Grüning (2007) does not find size to be directly significant in a structural equation model of disclosure causes, but nor does he allow it to be a direct cause – the model does not link size directly to disclosure at any point.

The above gives evidence that size is positively linked to disclosure. Many of the reasons given contain some implicit theories about disclosure, e.g. a theory that economies of scale occur. In addition, there are some possible effects from the studied theories. The PCH explicitly mentions that larger companies are more likely to receive negative attention. There is a possible Agency Theory explanation as it becomes harder to closely observe all managers as the scale of the firm rises, making agency problems more likely and disclosure a potential means to limit them.

The measurements of size are mostly drawn from related literature. Measures of a company's book value of assets are very common (e.g. Singhvi and Desai 1971, Cooke 1989, Raffournier 1995, Taylor et al 2010), although in many cases a logarithmic transformation is used. Few give any explanation for using this particular size measure; Malone et al (1993), focusing on oil and gas firms, state that it is one of the few measures available that is not affected by the fluctuations in oil prices. In DataStream, this is Total Assets (code wc02999).

A smaller number of papers use the companies' market capitalisation (e.g. Gelb and Strawser 2001, Ali et al 2007). This is a more accurate measure for companies that rely on intangible assets that cannot be recorded on the balance sheet, but is more subject to fluctuations in commodity prices than the more consistent book value. The gathered data is titled in DataStream as market Capitalisation (code wc08001).

The company's turnover is a common measure of size (e.g. Patten 1992, Lopes and Rodrigues 2007). Meek et al (1995) justify its use over other measures in an international study, arguing that it is less subject to variations in local GAAP than others. Additionally, it is another measure useful when the company relies on intangibles and would appear smaller under its book value. Revenue information is gathered using the data point titled Net Sales or Revenues in DataStream (code wc01001).

One paper using shareholders' equity has been identified (Grüning, 2007), although no justification for this measure is given. In practice, it offers the benefits of market value without being as subject to market fluctuations, although it does not measure size as directly. It is gathered in DataStream as Common Shareholders' Equity (code wc03501).

Cooke (1992) uses several size measures, including both some of those listed above and a number of alternative measures less common in the literature, arguing that each variable some unique information on company size that the others do not. Potential multicollinearity problems are prevented by merging these variables into a single size variable through factor analysis. This is conceptually similar to the use of latent variables in this thesis, using multiple measures to form a construct and using that to test models.

A measure of the company's total income is employed, using the datatype Operating Income (code wc01250). In addition, a measure of total profit was gathered, using Earnings Before Interest and Taxes (code wc18191).

5.2.2: Multiple Listing

Multiple listing status is widely thought to lead to greater disclosure. The common reasoning is that different countries have their own listing requirements that will not always overlap. For ease and cost reasons, multiple listed companies are likely to produce only one report that serves all requirements. This creates an unintentional voluntary disclosure; in each listing location, there is some information not required by the regulator as the company is working to another regulator's rules on what information must be provided. Saudagaran and Biddle (1995) study the decisions managers make about where to list when looking outside of their home location, arguing and finding supporting evidence that they tend to select areas that add the least to their existing disclosure requirements. If true, the effects of multiple listing on disclosure may be weaker than expected.

Some of the arguments are similar to those presented for size in section 5.2.1. A company listing in multiple locations often physically operates in each of them, meaning there are operations geographically removed from the headquarters. These will often be reported in a summarised report that is relatively easy to turn into an external publication. In addition, the facilities distant from the headquarters may have lower oversight and possible language barriers that combine to make information more easily hidden, creating a source of agency problems and making the theory applicable. Cooke (1989) states that multiple listing is a means of reducing cost of capital, as it allows shares to be issued in the markets of lowest cost.

If multiple listing is found to have effects on disclosure, this may be explained by Signalling Theory. Operating in geographically diverse areas may act as a form of diversification, making the firm more stable and therefore a better investment than comparable companies that are not as physically widespread.

In practice, literature has found multiple listing to have broadly positive effects on disclosure (e.g. Cooke 1989, Meek et al 1995, Robb et al 2001, Archambault and Archambault 2003, Abraham and Cox 2007, Grüning 2007, Lopes and Rodrigues 2007). Taylor et al (2010) offer a notable exception in which a negative effect is found.

Two measures are taken for multiple listing. Generally, researchers use a dummy variable for multiple listing status, equal to 1 if a company is listed in more than one country and 0 otherwise (e.g. Cooke 1989, Robb et al 2001). Abraham and Cox (2007) use a variation more relevant to this thesis: for their UK-based sample, multiple listing is only considered where the company is additionally listed in the USA, considered likely to increase disclosure due to increasingly strict rules being placed on USA listed companies. A US-listing dummy is adopted here.

The other measure is more wide-ranging and based on Saudagaran and Biddle's work (1995). As stated above, multiple listing should subject a company to further sets of rules and, as a result, lead to greater disclosure. To cover this effect, a measure of the total number of listings is used, counting the number of countries in which a company is listed other than the UK.

Neither measure was found as a coded variable within DataStream. They are instead obtained from company summaries within the program.

5.2.3: Financial performance

Recent financial performance is included as a variable. Arguments regarding its effect on disclosure work in both directions, however. One argument holds that low performance increases disclosure because managers need to show how they plan to resolve the situation (Armitage and Marston, 2008). The alternative is that managers become more open with good results as this means they can show their talents, while poor results lead to a comparative silence to prevent blame being placed on specific individuals (Singhvi and Desai 1971, Raffournier 1995).

Under some circumstances, it is possible that a low return may be a side effect of low disclosure. Underperformance in one division of the company can be obscured by pooling its returns with others when calculating this type of ratio. This means lowering disclosure as divisional information is not revealed, and has the effect of lowering the company's overall performance ratio.

Signalling Theory offers a very clear expected result for performance. Strong performance is a clear sign of good news that the company can and should highlight in order to stand out.

The literature examined has not produced a clear effect, supporting both of the opposing predictions above. While papers generally include positive predictions, the single most likely outcome is an insignificant effect on disclosure (e.g. Meek et al 1995, Watson et al 2002, Linsley et al 2006), with no clear uniting factor in the papers obtaining this result. Negative effects show some consistency, with two (Baginski et al 2002 and Chen et al 2002) using samples from the early- to mid-1990s, and the other two examples (Ali et al 2007, Chen et al 2008) each focusing on family-owned companies. The expected positive effects are the rarest of the three outcomes observed here, although negative results are only marginally more common (Singhvi and Desai 1971, Clarkson et al 1999, Field et al 2003). The total of positive and negative results is very close to the number of insignificant results for this variable.

Three measures of performance are used. The first is the return on assets (ROA), defined as Income/Book value of assets. This is a measure Singhvi and Desai (1971) used, and it has continued to be employed since (e.g. Taylor et al 2010). However, as stated under section 5.1.1, book value may not be the most reliable measure of size for all companies and therefore not the ideal means of scaling profits by size. Accordingly, additional measures are employed. Singhvi and Desai (1971) also used the earnings margin (EM), defined as Profit/Revenue, arguing that it better reflects the ability to absorb rising costs (the measure is used in some later papers, such as Meek et al 1995).

In addition, a third measure is used, that of Profit/Size, size measured as book value of assets, although there is little literature using this measure in disclosure.

None of these are obtained using DataStream directly, but each is calculated from two measures found in the program. Book value of assets, the denominator in ROA and Profit/Size, is as listed under 5.1.1 above, Total Assets (code wc02999). Income, used in ROA, is Operating Income (code wc01250). Profit, as used in EM and Profit/Size, is Earnings Before Interest and Taxes (code wc18191).

5.2.4: Debt finance

As the level of debt in a firm's capital structure increases, lenders become more important to the firm relative to shareholders. This has a possible effect on disclosure. Lenders, generally being large institutional investors, can often access information by pressuring managers. Shareholders, more likely to be individuals, lack the same ease of access and are more reliant on published information. With high levels of debt, disclosure becomes less important to the company as shareholders are relatively less important as a source of finance.

However, Agency Theory predicts that companies more likely to have agency problems would disclose more as a means of monitoring. Shareholders, being less powerful on an individual basis, are less important to the company than the lending institutions and the firm may start to operate for their benefit over the shareholders. This is an agency problem as some owners are not being served by the managers, so should lead to greater disclosure.

Perhaps reflective of the two strong yet contradictory lines of reasoning, the literature finds a mix of results on the matter. As in the financial performance case in section 5.1.3 above, the most likely result is an insignificant effect (e.g. Raffournier 1995, Watson et al 2002, Dobler et al 2010). Positive and negative effects are almost equal in number (and, combined, equal to insignificant results). There is little to link the positive results together (Malone et al 1993, Frankel et al 1995, Ahmed and Courtis 1999), and similarly little evidence to connect the negatives (Meek et al 1995, Zarzeski 1996, Brown and Hillegeist 2007, Taylor et al 2010).

Three measures are taken. Debt/Assets and Debt/Equity are both common measures used by many researchers (e.g. Meek et al 1995 and Ali et al 2007, respectively). Again, total assets may not reflect a given company's size well and debt/assets may not best reflect its proportion of debt, leading to the use of the second measure of debt/equity.

In addition, the total value of debt, unscaled by size, is collected and used as a measure of debt finance. However, it is questionable as a measure of the firm's capital structure. As shown below in tables 5.2 and 5.3, this measure correlates far more powerfully with size than other debt measures. This is not surprising; debt value is not scaled by size at all, yet a larger company may have obtained its size in part by including some debt in its financing to become larger than would be possible if relying on equity alone.

Like the performance measures, the two ratios are calculated using data from DataStream. Debt is measured using Total Debt (code wc03255) and divided by Total Assets (code wc02999, as above) for Debt/Assets and Common Shareholders' Equity (code wc03501) for Debt/Equity.

5.2.5: Sensitive industry

Many papers have included an industry membership variable as a determinant of disclosure. The reasons vary with the exact classifications used. There is reason to believe each industry forms its own disclosure practices that may vary from others (e.g. Botosan and Harris, 2000, and Patten, 1992). Some businesses will generate more information, for example, while follow-the-leader effects may result in comparable companies reaching similar practices. Legitimacy Theory holds that, as disclosure is a possible means to communicate, it is more important to some industries than others.

No consistently used industry classification variable is apparent in the literature. Researchers use different means of splitting their data depending on their exact research question and the sample used. Generally, any industries that are expected to disclose more on average are found to do so. Industry classifications found to disclose above average include resource extraction (Meek et al, 1995), chemicals (Robb et al, 2001), financial services (Lopes and Rodrigues, 2007), and retailers (Field et al, 2003).

Adams et al (1998) use four industry classes chosen due to their different characteristics. Two of these are found to disclose more than average. The first is oil, chemicals, metal, and power; a group of companies that exploit natural resources and process raw materials. The second is services, food, and retail; the group of companies that directly face the final consumer. This is broadly consistent with the industry classification findings of other papers and forms the basis for the industry variable employed for this thesis.

The industry variable used in this thesis is a dummy variable equal to 1 for 'sensitive' firms and 0 otherwise. Following the Legitimacy argument, and the example of Adams et al (1998), two kinds of firms are considered sensitive. The first are those that extract resources, included because they have clear environmental impacts that may harm the local area and its inhabitants and, in keeping with the theory, may suffer a loss of support from local stakeholders that leads to actions against the company. The other type is those which deal with the public directly. If a firm that deals with the general public attracts negative attention, the public can penalise it relatively easily by turning to competitors. A company further from the public in the supply chain is harder to influence in this way.

This variable is not gathered from DataStream and instead uses the FTSE website's industry classifications to define each company. Companies classed as some form of resource extraction, e.g. oil and gas, are considered sensitive, as are those with classes such as restaurants or retailers that deal with the general public. In practice, the industry classes considered sensitive are: Oil and gas production (0001 range); Chemicals (1300 range); Consumer goods (the entire broad 3000 range); Pharmaceuticals (4577); Consumer Services (5000 range, but excluding the narrow media agencies category 5555); Telecommunications services (6000 range); Utility providers (7000 range); Banks (8300 range); Forestry and paper (1730 range); Property and casualty insurance (8536); Life insurance (8570); Financial services (8770). Any classification not included in this list is considered non-sensitive.

The split of companies into expected high and low disclosure groups along these lines is broadly consistent with findings in the literature to date. The industries that have been observed to disclose more than most are those for which Legitimacy Theory provides an explanation for higher disclosure.

5.2.6: Return Volatility

The return volatility of a firm over time is included for several reasons. Arguments are mixed for this variable. One holds that the market rewards stability, which Signalling Theory predicts would lead to greater disclosure. The other is that volatile companies need to say more to the market since previous prices have lower predictive power.

Research is mixed on this variable. Insignificant effects (Gelb and Strawser 2001, Chen et al 2008) are rare, with a clear effect generally being visible when return volatility is used. However, the direction of effect varies, with some finding positive effects (e.g. Clarkson et al 1994, Abraham and Cox 2007) and others negative (e.g. Field et al 2003, Brown and Hillegeist 2007) with no clear pattern. This likely reflects the literature on volatility as a disclosure determinant, as competing explanations are given. It is possible that managers act in different ways depending on the company's circumstances.

The concept of volatility takes several forms as a company may vary in many ways from one period to the next. The major two are price-based volatility, i.e. changes to the share price, and returns volatility, i.e. changes in the level of profitability. Examination of the literature surrounding disclosure shows that papers making use of volatility tend towards return volatility, so the same is used here.

Only one variable is used for volatility. Return on equity values are collected over the past five years for each company. Where a company's data does no go back this far due to first listing at some point in the period (or first entering the database for any other reason), the data is used as far back as it exists. The standard deviation of the fluctuating return on equity is used as the actual variable (following the example of Field et al, 2002, and Chen et al, 2002). Data for this variable was readily obtained in DataStream using the return on equity datatype (code wc08301).

5.2.7: Disclosure

The measurement of disclosure in empirical studies of the concept is a challenge. Before a measurement can be selected, the researcher must first determine whether they will use disclosure quality or quantity. Two papers, Botosan (2004) and Beattie et al (2004), argue that disclosure quantity measures are not disclosure quality measures. While this seems simple, both papers point to others that have stated an intention to measure quality while using quantity measurements. In practice, few measurements will cover only one of these aspects exclusively, but in general quantity is far easier to measure and tends to dominate the research. Further, there is not a clear means of measuring disclosure quality. Regardless of whether quality or quantity is involved, papers tend to use one of three forms of measurement, discussed below.

The first method is to form an index of items of information expected and analyse a company's documents for them, scoring based on whether a company discloses the item or not. Singhvi and Desai (1971), one of the early papers in the area of disclosure studies, used this approach to measuring disclosure. The principle was soon discussed in more detail with Buzby (1974) discussing in detail which items should be included in such an index. The idea has since become a popular method of measuring disclosure, having been used consistently over time (e.g. Cooke 1989a, Robb et al 2001, Dobler et al 2011), although Rowley and Berman (2000) warn that an index approach may miss some details of interest to the researcher where a company scores well overall but is lacking disclosures in some areas.

There are a number of further negative points to this approach. The coding is subjective to some extent. While in many cases the presence or absence of an item will be apparent, there will be examples in which the wording of part of the document will make it unclear whether an item is covered or not. Further, something appearing in an unexpected location may be missed by the reader. This can be resolved to some degree by using multiple scorers, allowing two or more people to independently rate companies and comparing the results. Second, the use of a simple present/absent classification fails to capture the difference between those companies that give an item a passing mention and those that give it detailed analysis. Beattie et al (2004) suggest a solution to this latter problem with a more complex coding system (e.g. 1 point for mentioning an item and 2 for discussing it in depth), which adds complexity to the system and worsens the subjectivity problem.

The main concern for this project, however, is the intention to use a large sample size. For a smaller study, the detailed examination of documents for each company is practical. In this case, the sample is too large for this to be done in a practical timeframe.

The second measurement approach is to use an existing score. Various organisations rate company disclosures and provide lists of scores. The FAF/AIMR reports are a good example of this, having been used in various papers over time (e.g. Gelb and Strawser 2001, Welker 1995). The organisation behind these reports had groups of analysts specialised in various industry sectors examine company reports in great detail, checking for the presence of items of information that they would find useful to their own analysis, and formed a rating for each company from the combined score of each analyst who examined the company. In effect, this approach creates a formalised version of the disclosure scoring approach above, using very similar methodology with groups of expert analysts. It is therefore prone to the same problems of subjectivity and simple scoring as discussed above, although the use of multiple analysts to form a score ensures that the final ratings are a consensus. However, the researcher does not need to perform the scoring, eliminating the time-consuming part of the process.

The main problem with this approach is one of access to the resulting ratings. The FAF/AIMR reports were once commonly used but have not been made public since 1997. Further, using pre-existing scores limits the possible sample of companies to those included in the original scoring system. Continuing the FAF/AIMR report example, these reports only covered US-listed companies, preventing their use in any other national context.

The third approach is to find an indirect proxy measurement for disclosure. The two approaches above are also proxy measurements, but are based on disclosure activity directly; an indirect proxy, by contrast, uses the observation of a related matter to determine a company's disclosure. This is the method selected here as it has neither the time requirements of studying documents for the entire sample nor the access problems inherent to finding a pre-made disclosure score. The problem with this approach is that there is never a perfect proxy, regardless of the situation, as the proxy is not identical to the underlying measurement of interest.

Research using the proxy approach to disclosure measurement often uses properties of analyst forecasts, commonly one or more of the total analyst following (e.g. Lang and Lundholm, 1996), the forecast dispersion (e.g. Irani and Karamanou 2003), and less often the accuracy of forecasts (e.g. Francis et al 2006) or the scale of revisions made to forecasts (e.g. Lang and Lundholm, 1993). The use of any proxy based on analyst forecasts relies on an understanding of the role analysts play in the market, which must be discussed first. The question is whether analysts act as information intermediaries or information sources regarding the companies they analyse. In the former case, the analyst benefits from corporate disclosures and the number of analysts following the company should rise with disclosure. In the latter case, analysts compete with the information a company releases and greater disclosure is expected to lead to a smaller analyst following.

Lang and Lundholm (1996) performed an analysis primarily concerned with this matter. Their paper explains what observations would indicate analysts acting as sources or intermediaries and finds far more support for the information intermediary explanation. The finding is based on how analysts react to company-provided disclosures. Other information sources would compete with the company's own disclosures while intermediaries benefit from them. The analysts' role is therefore observable from their response to company-provided disclosure. If they are intermediaries then they will tend to analyse companies with strong disclosure policies, leading to more analysts following the company. If analysts instead act as information sources then they will compete with the company's disclosures, resulting in lower analyst followings where disclosure is stronger. Lang and Lundholm's (1996b) finding is the former; analyst following tends to be larger where the company discloses more. The paper discusses the possibility that the cause and effect is reversed, that the company's analyst following being large means there is more pressure on the company to release more information, but this interpretation does not change the overall finding that analysts act as information intermediaries; they are still demonstrating a preference for more disclosure provided by the company.

Broader evidence suggests that analysts take a more complicated role. Holland (1998) examines the results of disclosing information privately to selected analysts, observing that this practice helps to limit market responses, preventing extreme reactions to news. This suggests that analysts have a role in the market other than acting as an intermediary under these circumstances, likely acting as an information source in this case as they use information that is not otherwise available. Taken from the perspective of market participants without this privileged access to information, the analyst is acting as a source of information as it has not been revealed by the company directly. However, this practice is far less common now than it was at the time of publication due to the passing of Regulation Fair Disclosure in the USA and its influence on legislation elsewhere in the world and disclosure practices among companies. More recently, and not affected by this change, Ryan and Taffler (2004) observe evidence that analysts sometimes cause price movements, indicating that they are information sources to some extent. Ascioglu et al (2005) offer no analysis of the matter but assume analysts act as information sources; analyst following is used as a measurement of disclosure about a company that is not provided by the company in question. The evidence overall suggests that analysts have a complex role in the market. They act primarily as information intermediaries, but can act as information sources where circumstances permit.

There are a number of papers that use either analyst following or some aspect of their forecasts as a proxy measure of disclosure, usually taking (if not stating) analysts to act as intermediaries and benefit from the presence of company-provided information. Four measures based on analysts are common: the revisions made to analyst forecasts before results are published, the dispersion of analyst forecasts, the accuracy of analyst forecasts, and the total number of analysts following a company, each discussed below.

Revisions-based measures are uncommon in research. Lang and Lundholm (1993) find that revisions become less numerous and are of a smaller magnitude when the company's disclosure rating is higher. However, the expected effects on this variable are not entirely clear *a priori* as revisions to forecasts suggest analysts are surprised by new information. Such surprise could be the result of them learning of something not disclosed by the company, but equally could be the result of the company releasing information as soon as its employees are aware.

The accuracy of analyst forecasts compared to the actual result is often used to examine the effects of Regulation FD on financial analysts (e.g. Francis et al 2006, Heflin et al 2003). In this context, the variable is not directly a proxy for disclosure. However, Regulation FD represents a change in disclosure environment, intended to increase the overall information availability surrounding each company by preventing private disclosures. As a proxy for disclosure, it is expected that greater accuracy results from greater disclosure (as observed in Baik et al 2008). However, the result of using this measurement is dependent on the timing of forecasts relative to the publication of results. A forecast made (or revised) closer to the publication date is more likely to include all relevant information than one made far from publication.

Forecast dispersion is used with the assumption that detailed disclosure means all analysts tend towards the same conclusions (Lang and Lundholm 1996b, Irani and Karamanou 2003, Ali et al 2007). Different personal information acquisition levels, different analysis techniques, and differing capabilities among analysts all mean that analysts will draw non-uniform conclusions, but greater information availability results in a similarity in the core analysis. Dispersion is therefore a measure of analyst disagreement and is expected to fall with disclosure.

The final measure, analyst following, is more common in practice. Bhushan (1989) and Lang and Lundholm (1993) argue that analyst following should increase with a company's disclosure rating. Evidence generally supports this interpretation as analyst following rises with disclosure (e.g. Lang and Lundholm 1996b, Botosan 1997, Botosan and Harris 2000, Lang et al 2003, Ali et al 2007, Bozzolan et al 2009, Lang et al 2012). Healy et al (1999) examine the changes surrounding companies in periods in which their disclosure ratings improve. One of the observed effects is that analyst following rises after disclosure ratings rise, suggesting that analysts are inclined to follow companies with better disclosure ratings. A correlation between disclosure and analyst following exists.

Any combination of these four measurements can be used and a pilot study performed in the course of this thesis used all four. However, for the final study only analyst following was used. The reasons for the removal of the other three variables are largely due to differences in the two samples and the learning process from the pilot study. In the pilot study sample of the FTSE 350, most companies are followed by multiple analysts. The further the sample is expanded beyond this group, the more companies have a following of only a single analyst or none at all. In either case, analyst dispersion becomes meaningless; there is zero dispersion regardless of the company's disclosure policy. The number of forecast revisions is similarly problematic in cases with zero following as there can be no revisions by default. This measure has further problems with low total followings as the total number of revisions is limited by the total number of estimates. Similarly, with no analysts there is no potential to identify the accuracy of forecasts, and small followings lead to a few estimates dominating the calculated accuracy value. Total analyst following is not subject to these problems as a low following simply means a low value is recorded for that company and there are no issues with the measurement.

Bhushan (1989) and Lang and Lundholm (1993) each explain in part how and why analyst following responds to disclosure. Each analyst or organisation employing them is a business and, for the purposes of this argument, they are considered to be rational wealth maximisers, in keeping with the conditions of the theories laid out in section 4.7, so analysts are driven by profit.

The business process for analysts as information intermediaries is to acquire information, process it in some way (likely drawing conclusions about the company's prospects), and sell the resulting package of information that includes both the company's own release and the analyst's additions. Information economics mean the costs of production scale little with the number of sales; once the analysis process is complete, the results can be sold to multiple parties with only minimal further production costs (for example, printing the information). Bhushan (1989) mentions that this may also work against the analyst as a buyer may be able to resell the information.

As in most business models, one of the approaches available to analysts to maximise profit is by lowering costs. The very low per-unit costs mean that the primary source of cost reduction is in the information gathering and processing stages. The cheapest companies to follow, then, are those for which information is plentiful and easily processed. This trait depends largely on the company's own information releases, i.e. its disclosure policy. Detailed disclosures make information cheap for analysts to acquire, although the potential to obscure details in a mass of information means it is not necessarily cheap to process. High-quality disclosure should solve this problem by releasing relevant information and making for ease of processing. For the analyst, the ideal is a company that releases both detailed and high-quality information as this is cheap to both acquire and analyse.

The result of this discussion is that analysts are more likely to follow those companies that produce detailed or high quality disclosures, and much more likely to follow those that produce both.

For the research performed here, the analyst following proxy was obtained from DataStream. The number of EPS estimates for the next financial year (coded as F1NE) was used. Although this is not necessarily the actual number of analysts taking an interest in the company, it shows the number following closely enough to be confident in making a prediction of results in the near future.

The use of analyst following has another benefit. As discussed above, the alternative approaches are to use an existing disclosure score or form an index. Existing scores are often formed with reference to analysts, making the resulting ratings indicative of what analysts want to see in a company's disclosures. Similarly, when making an index the researcher needs to select items to include and this often involves seeking analysts' views on what makes a useful disclosure. Analyst following is therefore related to these alternative approaches as by its nature it captures analyst preferences for what is disclosed.

There is a notable and important disadvantage to analyst following as a measure of disclosure, however. As mentioned in section 2.2, compliance with mandatory disclosure requirements will vary among companies as some managers will comply with the letter of the rules and others the spirit. Analyst following is argued here to respond to the overall disclosure quality or quantity, regardless of whether the disclosures in question are required or voluntarily provided. While the measure can be reasonably expected to primarily reflect voluntary disclosure, it does not do so exclusively and some amount of variation within mandatory disclosure will be contained within this proxy.

5.3: Data properties

This section covers the results of descriptive analysis performed on each variable along with correlation analysis on all of them, collectively. Tables of overall descriptive statistics and correlations are presented first, then discussed in detail in the same order as in Section 5.2. Discussion here focuses mainly on the basic untransformed data set with all observations included. Generally, each transformation results in stronger correlations than the basic data, but very rarely shows a significant relationship where the basic data did not already identify one.

Table 5.1 below presents the descriptive statistics for each of the variables. This covers the minimum, maximum, mean, standard deviation, skewness and kurtosis for all of the non-binary variables.

Tables 5.2 and 5.3 are each tables of correlations among the variables used in models later in the thesis. Table 5.2 covers the Pearson correlation, the common interpretation of correlation that tests for a linear relationship between two variables. Table 5.3 instead uses the Spearman correlation, a non-parametric (and less commonly used) measurement of correlation. The Spearman correlation uses the rankings of observations rather than their actual values to obtain a correlation value, recording high values if the relationship between two variables is monotonic but not necessarily linear. A curvilinear relationship that is perfectly monotonic would have a Spearman correlation of 1 but a lower Pearson correlation due to the nonlinear link between variables. While useful for identifying that a relationship exists, the Spearman correlation is poor for defining a meaningful strength. Cooke (1998) uses a good example to explain the problem. A Spearman correlation of 0.7 indicates that an increase of 1 to one variable is expected to increase the other by 0.7. However, both variables are ranks and therefore take integer values; a change of 0.7 is not possible. Interpretation of any values other than 1, 0, or -1 has the same problem. The strength of a relationship can be stated relative to other Spearman correlation values (e.g. 0.5 indicating a stronger link than 0.3), but the use of ranks means the power cannot be interpreted in terms of the units of the variables in question.

The use of both correlations gives a better overall understanding of how the variables are related since it is easily possible for two variables to share a very strong relationship but have a low correlation value using only one technique.

Table 5.1: Descriptive Statistics

	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Size measures						
Book Value (£000)	0.02	1,651,255.32	7,570.21	78,504.64	16.44	295.30
Market Cap (£000)	0.00	1,518,667.85	2,925.92	41,914.74	33.37	1195.72
Profit (£000)	-2,356.00	344,600.00	538.05	10,453.39	29.06	893.71
Revenue (£000)	-9.45	425,090.30	1,925.46	16,168.71	19.33	433.67
Income (£000)	-5,096.00	177,180.45	297.62	4,825.10	34.58	1261.59
Equity (£000)	-286.50	629,538.65	1,416.20	17,800.60	31.22	1084.78
Foreign listing measure						
Foreign listing (# of countries)	0.00	6.00	0.81	0.78	.84	1.50
Debt finance measures						
Debt (£000)	0.00	439,559.00	1,319.48	16,154.41	21.14	497.18
Debt/ Equity (£000)	-64.68	171.40	0.38	5.49	19.12	677.70
Debt/ Assets (£000)	0.00	32.42	0.20	1.02	26.30	768.14
Performance measures						
Profit/Size	-31.70	144.43	0.01	4.09	31.03	1095.28
Earnings Margin	-314,179.00	312,499.00	-458.22	13,346.52	-2.59	453.64
Return on Assets	-11.06	3.02	-0.08	0.63	-9.07	121.24
Volatility measure						
Volatility	0.00	3383.29	41.10	165.08	13.99	238.59
Disclosure measure						
Following (# analyst estimates)	0.00	32.00	4.35	6.67	1.97	3.14

Overall, table 5.1 indicates that most of the variables display high kurtosis. While less obviously problematic, skewness is also high on most variables. The one variable that comes closest to the desired normal distribution according to the table is multiple listing, an unexpected result for a variable that can take a range of values limited to whole numbers between 0 and 6.

At a glance, the financial performance measures appear incorrect. Each is based on the overall profitability measure scaled by some other measurement. Profit/Size, Earnings Margin and Return on Assets each take a mean value that appears unreasonably low. However, examining the data, these values are expected. Profit values in the sample are often low, with 575 companies reporting a loss at the time of data collection. Profit/Size has a low mean of 0.01 due to the effects of so many negative profit reports. The negative mean observed for Earnings Margin is due to a stronger version of the same effect. 147 companies reported zero or negative revenue values (which likely explains to the negative profit reported in all such cases). Attempting to use these values in the Profit/Revenue formula of EM would create further problems. Where revenue is reported as 0, the formula would divide by zero and create invalid results. In the few cases where the reported revenue is negative, there is a larger problem. In all of these, the profit value is (as should be expected) also negative. The formula would then divide a negative by a negative. While mathematically possible, the result would be a positive EM in cases where the company's performance is very poor, and possibly even a large EM where the profit is orders of magnitude lower than the revenue. To minimise the effects of both the negative and zero revenues, the EM formula is altered to instead divide by 1, leading to results equal to the reported profit. The Return on Assets values have a similar cause, although less pronounced as total reported income tends to be less extreme than reported profit whether a profit or loss is reported.

In addition, a brief examination of the companies reporting negative performance ratios was carried out using data from later years. Most of the companies in this situation either improved their performance rapidly or were de-listed within two years.

The remaining parts of section 5.3 discuss tables 5.1, 5.2, and 5.3. Each variable is discussed in terms of the observed correlations with others and expectations are formed for the upcoming analysis in chapters 6 and 7. However, as these correlation analyses are univariate tests and the models are multivariate, some differences between expectations and observations are likely.

Table 5.2: Pearson Correlations

	Profit	Book Value	Market Cap	Revenue	Income	Equity	Foreign listing	US listing	Profit/Size	Earnings Margin	Return on Assets	Debt	Debt/ Equity	Debt/ Assets	Earnings Volatility	Sensitive	Analyst Following
Profit	1.000	.175**	.488**	.393**	.493**	.478**	.048	.025	.007	.004	.020	.070**	-.002	-.006	-.010	-.019	.118**
Book Value	.175**	1.000	.367**	.427**	.311**	.498**	.175**	.131**	.009	.005	.018	.898**	.046	-.003	-.018	.078**	.273**
Market Cap	.488**	.367**	1000	.783**	.989**	.968**	.105**	.087**	.012	.005	.026	.161**	.006	-.003	-.012	.013	.158**
Revenue	.393**	.427**	.783**	1000	.796**	.848**	.167**	.107**	.015	.006	.034	.248**	.007	-.002	-.018	.065*	.282**
Income	.493**	.311**	.989**	.796**	1000	.965**	.090**	.072**	.011	.004	.025	.102**	-.003	-.003	-.010	.008	.137**
Equity	.478**	.498**	.968**	.848**	.965**	1000	.121**	.098**	.011	.006	.024	.288**	.008	-.004	-.015	.023	.182**
Foreign listing	.048	.175**	.105**	.167**	.090**	.121**	1000	.636**	.008	-.067*	.054*	.142**	.015	-.044	-..047	.234**	.403**
US listing	.025	.131**	.087**	.107**	.072**	.098**	.636**	1000	.012	-.070**	.044	.116**	-.019	-.033	-.018	.095**	.189**
Profit/ Size	.007	.009	.012	.015	.011	.011	.008	.012	1000	.029	.376**	.007	-.002	-.109**	.007	-.019	.069**
Earnings Margin	.004	.005	.005	.006	.004	.006	-.067*	-.070**	.029	1000	.027	.004	.002	.006	-.011	-.083**	-.010
Return on Assets	.020	.018	.026	.034	.025	.024	.054*	.044	.376**	.027	1000	.014	.016	-.328**	-.103**	-.017	.162**
Debt	.070**	.898**	.161**	.248**	.102**	.288**	.142**	.116**	.007	.004	.014	1000	.075**	.005	-.014	.076**	.239**
Debt/ Equity	-.002	.046	.006	.007	-.003	.008	.015	-.019	-.002	.002	.016	.075**	1000	.010	-.086**	.048	.010
Debt/ Assets	-.006	-.003	-.003	-.002	-.003	-.004	-.044	-.033	-.109**	.006	-.328**	.005	.010	1000	.008	-.014	-.006
Earnings Vol	-..010	-.018	-.012	-.018	-.010	-.015	-.047	-.018	.007	-.011	-.103**	-.014	-.086**	.008	1000	.023	-.046
Sensitive	-.019	.078**	.013	.065*	.008	.023	.234**	.095**	-.019	-.083**	-.017	.076**	.048	-.014	.023	1000	.169**
Analyst Following	.118**	.273**	.158**	.282**	.137**	.182**	.403**	.189**	.069**	-.010	.162**	.239**	.010	-.006	-.046	.169**	1000

*: Correlation significant at the 0.05 level (2-tailed)

**: Correlation significant at the 0.01 level (2-tailed)

Table 5.3: Spearman Nonparametric Correlations

	Profit	Book Value	Market Cap	Revenue	Income	Equity	Foreign listing	US listing	Profit/Size	Earnings Margin	Return on Assets	Debt	Debt/Equity	Debt/Assets	Earnings Volatility	Sensitive	Analyst Following
Profit	1.000	.581**	.526**	.689**	.839**	.537**	.069**	.049	.798**	.785**	.718**	.447**	.276**	.198**	-.305**	-.020	.553**
Book Value	.581**	1.000	.829**	.837**	.634**	.928**	.366**	.210**	.460**	.453**	.502**	.724**	.462**	.342**	-.247**	.120**	.739**
Market Cap	.526**	.829**	1.000	.698**	.559**	.812**	.363**	.214**	.426**	.387**	.463**	.526**	.283**	.165**	-.213**	.098**	.726**
Revenue	.689**	.837**	.698**	1.000	.791**	.726**	.190**	.110**	.579**	.533**	.658**	.711**	.493**	.390**	-.248**	-.006	.717**
Income	.839**	.634**	.559**	.791**	1.000	.570**	.119**	.052	.698**	.658**	.795**	.539**	.348**	.267**	-.281**	.003	.628**
Equity	.537**	.928**	.812**	.726**	.570**	1.000	.377**	.224**	.422**	.424**	.443**	.594**	.410**	.186**	-.278**	.134**	.707**
Foreign listing	.069**	.366**	.363**	.190**	.119**	.377**	1.000	.607**	-.001	-.011	.029	.200**	.067*	.017	.022	.223**	.340**
US listing	.049	.210**	.214**	.110**	.052	.224**	.607**	1.000	.029	.012	.036	.064*	-.010	-.045	-.033	.095**	.198**
Profit/ Size	.798**	.460**	.426**	.579**	.698**	.422**	-.001	.029	1.000	.859**	.897**	.307**	.217**	.148**	-.262**	-.105**	.437**
Earnings Margin	.785**	.453**	.387**	.533**	.658**	.424**	-.011	.012	.859**	1.000	.744**	.336**	.249**	.189**	-.255**	-.136**	.391**
Return on Assets	.718**	.502**	.463**	.658**	.795**	.443**	.029	.036	.897**	.744**	1.000	.357**	.260**	.187**	-.272**	-.097**	.499**
Debt	.447**	.724**	.526**	.711**	.539**	.594**	.200**	.064*	.307**	.336**	.357**	1.000	.783**	.822**	-.133**	.058*	.541**
Debt/ Equity	.276**	.462**	.283**	.493**	.348**	.410**	.067*	-.010	.217**	.249**	.260**	.783**	1.000	.763**	.015	.005	.309**
Debt/ Assets	.198**	.342**	.165**	.390**	.267**	.186**	.017	-.045	.148**	.189**	.187**	.822**	.763**	1.000	-.025	-.025	.224**
Earnings Vol	-.305**	-.247**	-.213**	-.248**	-.281**	-.278**	.022	-.033	-.262**	-.255**	-.272**	-.133**	.015	-.025	1.000	.050	-.213**
Sensitive	-.020	.120**	.098**	-.006	.003	.134**	.223**	.095**	-.105**	-.136**	-.097**	.058*	.005	-.025	.050	1.000	.120**
Analyst Following	.553**	.739**	.726**	.717**	.628**	.707**	.340**	.198**	.437**	.391**	.499**	.541**	.309**	.224**	-.213**	.120**	1.000

*: Correlation significant at the 0.05 level (2-tailed)

**: Correlation significant at the 0.01 level (2-tailed)

5.3.1: Size

The collected size variables – profit, book value of assets, market capitalisation, revenue, income, and equity – are all considered collectively due to (expected) similarities in correlations. As shown in table 5.1 above, these all display a high degree of non-normality, having very high kurtosis and large positive skewness. The skewness is to be expected by the nature of the variables, which for the most part have a lower bound of 0 and no practical upper bound.

Plotting the variables explains much of the kurtosis. Each variable displays at least one very large outlier and a number of others that, although smaller, are still above the typical range for the variable.

The size measures display a high degree of Pearson correlation with each other, as expected of alternative measurements of the same underlying concept. Among these six, book value and income are the least correlated, with only $\rho=0.175$. Of the six, these two generally demonstrate the lowest correlations with others in the group; the highest correlation profit has is $\rho=0.493$ with income, while book value's maximum is $\rho=0.498$ with equity. By contrast, the lowest correlation among the remaining four size measures is $\rho=0.783$ between market cap and revenue. Market cap, income, and equity all share strong correlations with each other, ranging from $\rho=0.965$ to $\rho=0.989$.

If Spearman's nonparametric correlation values are used instead of the Pearson correlations discussed above, the values of the correlations change. The results still show strong correlations within the group, but income and book value integrate better and are not noticeably less correlated than others. This suggests these two do not scale linearly with other size measures.

The size measures generally demonstrate correlations with three variables outside of the group. A correlation is observed between size measures and total foreign listing in all cases except for profit. The remaining five size measures all display a significant correlation with foreign listing, although not a large one (maximum is $\rho=0.175$ with book value). The same observation is made with size measures and US listing status, although in each case the correlation is lower. These relationships were expected; a company that has operations in multiple countries has more potential for growth than one restricted to a single area. Spearman correlations show stronger connections between these variables and the size measures but otherwise the observations are unchanged other than the profit and total listing correlation being significant, though small (ρ=0.069).

Pearson correlations suggest that there is no link between size and performance, with no significant correlations identified between any performance and any size variable. Spearman correlations in all cases are significant and positive, and of moderate to powerful strength (ρ=0.422 at the lowest, between equity and profit/size, and ρ=0.798 between profit and profit/size). This suggests that there is a notable monotonic correlation between size and performance, but not a linear one.

All size measures demonstrate some correlation with the total value of debt in the company, suggesting that debt is a more reliable measure of size than debt finance. While these correlations are generally low, the company's book value demonstrates a very high correlation with debt at ρ=0.898. This may be evidence that debt finance is used in addition to equity finance rather than acting as a substitute as it suggests that companies using debt are able to purchase more assets than those that do not use this alternative source of finance.

If size and debt ratios are correlated using Spearman's method then, like with performance, a new interpretation results. All debt measures demonstrate a significant positive correlation with size using this technique. However, the two debt ratio measures are never strongly correlated with size, at most reaching ρ=0.493 (debt/equity and revenue). The debt value measure obtains higher correlations but remains correlated with size measures. As with performance, this all suggests the relationship between size and debt finance is somewhat monotonic but not linear.

Earnings volatility is comparable to the performance measures in that it is not correlated with any size measure using Pearson correlations but has moderate Spearman correlations (ranging from ρ=-0.213 to ρ=-0.305). Again, the relationship is not linear.

Book value and revenue each show very low correlations with sensitive industry membership (ρ=0.078 and 0.065 respectively). It is not clear why these two variables should be connected to the industry type, although the low correlation values suggest it may be a property of the sample and not a general observation to expect in other cases. Spearman correlations do not change much with regard to sensitivity as, although the details differ, the result is still low correlations with some but not all size measures.

Finally, the size measures are all correlated with analyst following. The correlations range from ρ=0.118 (profit) to 0.282 (revenue). This provides some evidence that the sample used here demonstrates the commonly-observed link between company size and disclosure practices. The Spearman correlations are higher (reaching a maximum of ρ=0.739) but otherwise the overall observation of a link between size and analyst following holds. Once again, the relationship here appears to be largely monotonic but non-linear.

5.3.2: Multiple Listing

The measure for total foreign listing has a mean of 0.81, indicating a tendency towards few listings outside of the UK. When multiple listing is measured only by reference to US listing, little changes in the analysis. This measure does not appear in table 5.1 above as it is a binary variable and most of the measures taken there would not apply; the details are presented below in table 5.4. Approximately one fifth of the sample is listed in the USA.

Table 5.4: US listing

	Frequency	Percent
Not US listed	1161	80.8
US listed	275	19.2

The total foreign listing variable displays a high degree of normality compared to others in the sample, having low kurtosis and skewness values as seen in table 5.1. However, as a finite and countable variable taking only a very small range of values, the normal distribution is not strictly appropriate to the variable.

The two listing variables, foreign and US listing, display a correlation of ρ=0.636 (Spearman ρ=0.670). Total foreign listing includes US listing, which explains much of this correlation. As mentioned above, both variables have low but significant correlations with the size measures.

Both listing variables display a very low, significant negative Pearson correlation with earnings margin (ρ=-0.067 with total listing, -0.070 with US listing) but no significant correlations with other financial performance measurements. Total foreign listing displays a similar correlation (ρ=-0.055) with the volatility of recent earnings. Spearman correlations between listing and performance are uniformly insignificant, in keeping with the very low Pearson correlations.

Listing has significant Pearson correlations with only one of the debt finance variables, the actual value of debt. Spearman correlations change little, creating one additional significant correlation between debt/equity and foreign listing, but the value of this is very low (ρ=0.067).

Both variables display a significant correlation with sensitive industry membership. While the value of this correlation is low for US listing (ρ=0.095), the value for total listing (ρ=0.234) cannot be dismissed. Spearman correlation values are barely different.

Finally, both variables show a correlation with analyst following, taking values of ρ=0.403 for total listing and ρ=0.189 for US listing, and again the Spearman correlation values are similar. This supports the idea that listing in multiple countries will lead companies to disclose more information due to differing listing requirements in each regulatory jurisdiction. Note that the ρ=0.403 correlation with total listing is the highest Pearson correlation involving analyst following.

5.3.3: Financial Performance

All of the performance measures display high positive kurtosis, suggesting that a normal distribution cannot be applied. Unlike others in the sample, two of the performance variables have negative skewness (ROA and earnings margin). All three of the performance measures are constructed as a ratio of two of the size variables, a numerator related to the throughput and a denominator related to the recorded value of the firm, and the properties of these ratios are therefore likely to be influenced heavily by at least one of the variables involved. If plotted, these tend to have potential outliers at both ends of the scale due to a few companies recording high profits while others experience high losses. Visually, profit/size and earnings margin each display one very large outlier at the upper end (although not the same company in each case) and have a few outliers at the lower end of varying scale. ROA behaves differently, having a varied range of low outliers and a few smaller outliers above its typical range.

The three performance measures are not strongly correlated with each other. ROA and profit/size are correlated with $\rho=0.376$, while earnings margin is not correlated with others in the group. As seen in chapter 7, models that initially used all three invariably removed earnings margin as insignificant, suggesting that it is not compatible with the others. Spearman correlations show a different relationship, however, with all three having strong positive correlations (minimum $\rho=0.744$).

In addition to the correlations between performance measures and listing discussed in the relevant sections, there are a few other significant correlations. First, profit/size and ROA are each correlated with the debt/assets measure, taking values of $\rho=-0.109$ and $\rho=-0.328$ respectively. This suggests that companies have better performance when debt finance levels are low in relation to the recorded asset value. However, all three of these variables involve the use of book value of assets as a denominator, and the link may be related more to this than any accounting-based explanation. The Spearman correlations show a very different relationship between performance and debt finance, with low but uniformly positive correlations between any pair of measures.

Pearson correlations suggest no relationship between performance and earnings volatility. Spearman correlations suggest otherwise, having moderate correlations between performance measures and volatility (profit/size $\rho=-0.262$, earnings margin $\rho=-0.255$, and ROA $\rho=-0.272$).

Further evidence that earnings margin acts differently to the other performance measures is observed in its significant correlation with sensitive industry membership of $\rho=-0.083$. Spearman correlations, however, increase the magnitude of this relationship to ρ=-0.136, but significant negative correlations with the other two performance variables mean this is comparable to others.

In addition, profit/size and ROA (but not the earnings margin) each show correlations with the analyst following. The profit/size correlation is very low at ρ=0.069, while the ROA correlation is higher at ρ=0.162. Spearman correlations instead show a link of approximately ρ=0.400 with all three performance measures.

5.3.4: Debt Finance

The debt finance measurements each display similarly high positive skewness and kurtosis. The value of debt in the firm appears visually similar to a size variable, having a large number of outliers of varying sizes, while the two ratio measurements have a small number of large positive outliers.

There are very few correlations involving the measures of debt. The three measures in the group are barely correlated, with only debt and debt/equity showing anything significant, but the value is a very low ρ=0.075. Spearman correlations show a different relationship with all three having strong correlations (minimum ρ=0.763), suggesting a non-linear relationship between these variables.

As mentioned above, the size measures correlate with the total (i.e. unscaled) value of debt and listing status, much like the size variables under Pearson values while all three debt finance measures correlate to some extent under Spearman values. Listing status correlates only to debt value and Spearman correlations do not change much in this category. As discussed earlier, the performance measures correlate only when using Spearman's methods.

Earnings volatility has no significant Pearson correlation with any debt finance measure, but significant positive Spearman correlations occur. The debt value has ρ=0.350, comparable to size measures, while the two debt ratios have ρ=0.283 for debt/equity and ρ=0.201 for debt/assets, suggesting once again that there is a non-linear relationship between volatility and debt finance usage.

There is a very small correlation between debt value and sensitive industry membership (ρ=0.076), again comparable to some size measures. This does not change if Spearman correlations are used, with the debt and sensitivity correlation actually lower under this version of correlation (ρ=0.058).

The value of debt displays a significant correlation with analyst following in line with the size measurements (ρ=0.239), while the two measures of debt scaled by company size do not show significant correlations. This suggests that debt finance level is unlikely to have an effect on analyst following in later models and further supports that total debt is more indicative of firm size than the use of debt finance. Spearman correlations are similar to Pearson in this case, with the debt value correlation being ρ=-0.133 and no significant correlations with the other two.

5.3.5: Sensitive industry

As a binary measure, the sensitive industry variable is not described in table 5.1 above. Its frequencies are in table 5.5 below. As it shows, there is an almost even split between sensitive and non-sensitive industries in the sample.

Table 5.5: Sensitive Industry

	Frequency	Percent
Non-sensitive industry	771	53.7
Sensitive industry	665	46.3

Most correlations involving sensitive industry membership are mentioned in one of the sections above. It is significantly (Pearson) correlated with some size measures, listing status, and the value of debt to low extents and has a low negative correlation with earnings margin but no other performance measure. Spearman correlations are not very different for the most part other than being low and negative with all three performance measures.

Finally, it is worth noting that that sensitivity is correlated with analyst following. At ρ=0.169, its value is comparable to the correlations between following and size. This relationship is weaker (ρ=0.120) but still exists according to the Spearman correlation.

5.3.6: Return Volatility

The non-normality of the volatility measure is comparable to that of most of the size variables, showing similar skewness and kurtosis to many of them. When plotted, this variable is visually similar to most of the size measures as there are a few large outliers and a larger number of smaller ones.

There are only two significant Pearson correlations involving earnings volatility, the low (ρ=-0.086) correlation with debt/equity and a marginally higher (ρ=-0.103) correlation with return on assets. The variable is otherwise unrelated to any other under the Pearson method of correlation.

Spearman's non-parametric correlation creates a very different image. Under this method, significant correlations are present with a range of other variables. The lack of significant Pearson correlations combined with the number of significant Spearman correlations indicates that volatility has strong but highly non-linear relationships with other variables.

The six size measures each display a significant correlation ranging from ρ=-0.213 (market cap) to ρ=-0.305 (profit), suggesting that larger companies have more stable returns. Correlations of similar scale exist with the three performance measurements, suggesting that higher income tends to be linked with stable returns. The value of debt, which has been previously indicated to act more like a size measure than a debt finance measure, displays a lower correlation (ρ=-0.133).

In addition, a significant correlation of ρ=-0.213 exists with the disclosure measure, indicating a small tendency for companies with more stable returns to provide more information. This may be a case of the two variables having a common cause rather than a causal link; larger companies have more to disclose and are more stable, as indicated by correlations between size and both disclosure and volatility.

Unusually, there is not a significant Spearman correlation with the debt/equity measure despite there being a significant Pearson correlation between this and volatility.

5.3.7 Outliers

Outliers, observations that take unusually extreme values, are a possible problem with any data set. Regression analysis relies on a line of best fit, one that minimises the total distance between itself and all observations, as discussed in section 5.4.1. In OLS regression, all observations are equally weighted; the distance between the line and any one observation is as important to minimise as the distance to any other. This creates a clear problem if outliers are involved in the data set. A clear pattern may exist in most of the observations, but an outlier will tend to pull the line of best fit away from this pattern as it is considered equally important.

Other regression approaches may apply unequal weights to observations. This makes it more important to minimise the distance to some observations than others as the highly-weighted cases will contribute larger amounts to any distance measurement. Unequal weighting can reduce the problems caused by outliers by making their distances less important to the overall function, but it cannot completely eliminate the problems. It is also possible for an unequal weighting system to assign high weight to an outlier and further move the final line of best fit away from a pattern apparent in other variables.

The handling of outliers creates an issue for quantitative research. Including an outlier in the sample means the problems described above may occur. Removing them from the sample eliminates these problems, but means relevant data is unused. Outliers may be due to errors made in the process of data entry, e.g. typing an extra digit and increasing a value by an order of magnitude. Cases like this are safe to remove from a data set as they are incorrect and their effects on patterns are not reliable. Other cases are genuinely unusual observations that behave differently from others in the sample. Observations of this type may be included or removed from the sample. Their inclusion may make a pattern or relationship that exists elsewhere in the sample less clear. However, their removal means not including true and valid observations. It is also often difficult to determine which outliers are errors and which are genuine but unusual observations.

The data used in this thesis come from Datastream outputs, which rely on data entered into Datastream, which in turn are provided by the company. This means errors may occur at the company level, at the Datastream level, or at the researcher level. Outliers are checked against the company's published reports where possible, which eliminates Datastream and researcher-caused error. Company-level errors are still possible, although for listed companies the information should have been entered carefully due to possible legal action and audited before publication, largely eliminating material errors. It is therefore unlikely that outliers in the data set are due to error and most are accurate but extreme values.

Outliers in this thesis are investigated using the Mahalanobis Distance method. This approach helps to identify multivariate outliers, observations that are unusually large or small in multiple variables once correlations in the sample are taken into account. The test converts the Mahalanobis distance into a p-value, then searches for cases where $p<0.001$. Any observation with a p-value in this range has a highly significant distance and is very likely to be an outlier. When this test is applied to the sample used in this thesis, 46 outliers are identified. Most of these companies have a clearly identifiable reason to be considered an outlier, such as a recorded loss meaning the performance ratios take extreme negative values, or cases where the company uses such extreme levels of debt in its capital structure that the debt-to-equity ratio takes a large value. A few companies record negative equity or a book value that exceeds the market value of the company. In the time since data was gathered, most of these companies have either changed their situation or ceased to exist as independent and active business entities.

Descriptive statistics and correlation tables for the outlier-free sample are presented below. Removing outliers reduces skewness and kurtosis and tends to strengthen correlations.

Table 5.6: Descriptive Statistics with Outliers Removed

	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Size measures						
Book Value (£000)	53	167011000	1671947.84	9590167.861	11.998	169.274
Market Cap (£000)	0	68453931	985894.51	5004660.135	10.144	116.658
Profit (£000)	-9452	41591430	781990.33	2953969.631	6.886	59.744
Revenue (£000)	-853320	12284123	101703.16	579678.432	12.266	193.608
Income (£000)	-286500	28821975	421588.21	1743881.674	9.261	108.047
Equity (£000)						
Foreign listing measure						
Foreign listing (# of countries)	0.00	6.00	0.81	0.78	.84	1.50
Debt finance measures						
Debt (£000)	0	29776000	302653.75	1522151.213	10.326	143.129
Debt/ Equity (£000)	-10.158	16.674	.401	1.436	2.587	40.567
Debt/ Assets (£000)	0.000	2.106	.156	.223	3.153	16.518
Performance measures						
Profit/Size	-2.874	.862	-.055	.323	-3.551	18.946
Earnings Margin	-33614.000	9697.000	-338.843	2077.269	-9.372	115.665
Return on Assets	-1.962	1.241	-.029	.263	-2.846	12.912
Volatility measure						
Volatility	0.000	637.435	28.278	55.361	5.477	40.198
Disclosure measure						
Following (# analyst estimates)	0.00	32.00	4.35	6.67	1.97	3.14

Table 5.7: Pearson Correlations with Outliers Removed

	Profit	Book Value	Market Cap	Revenue	Income	Equity	Foreign listing	US listing	Profit/Size	Earnings Margin	Return on Assets	Debt	Debt/ Equity	Debt/ Assets	Earnings Volatility	Sensitive	Analyst Following
Profit	1.000	.343**	.797**	.609**	.871**	.496**	.156**	.052	.124**	.028	.119**	.534**	.077**	.062*	-.038	.100**	.354**
Book Value	.343**	1.000	.433**	.519**	.470**	.549**	.223**	.073**	.056*	.026	.060*	.533**	.095**	.019	-.052	.097**	.335**
Market Cap	.797**	.433**	1.000	.770**	.929**	.790**	.209**	.048	.103**	0.016	.117**	.713**	.072**	.076**	-.048	.119**	.468**
Revenue	.609**	.519**	.770**	1.000	.754**	.755**	.234**	.070**	.113**	.043	.129**	.674**	.098**	.086**	-.065*	.139**	.551**
Income	.871**	.470**	.929**	.754**	1.000	.758**	.209**	.064*	.108**	0.031	.128**	.674**	.075**	.071**	-.045	.128**	.411**
Equity	.496**	.549**	.790**	.755**	.758**	1.000	.283**	.091**	.091**	.027	.105**	.702**	.053*	.056*	-.076**	.127**	.484**
Foreign listing	.156**	.223**	.209**	.234**	.209**	.283**	1.000	.629**	.009	-.132**	.034	.194**	.034	-.016	-.019	.213**	.366**
US listing	.052	.073**	.048	.070**	.064*	.091**	.629**	1.000	.016	-.136**	0.032	0.032	-0.032	-.060*	-.044	.075**	.153**
Profit/ Size	.124**	.056*	.103**	.113**	.108**	.091**	.009	.016	1.000	.180**	.864**	.075**	.088**	.003	-.276**	-.053	.256**
Earnings Margin	.028	.026	.016	.043	.031	.027	-.132**	-.136**	.180**	1.000	.134**	.033	.042	.078**	-.074**	-.133**	.021
Return on Assets	.119**	.060*	.117**	.129**	.128**	.105**	.034	.032	.864**	.134**	1.000	.085**	.096**	.022	-.280**	-.050	.282**
Debt	.534**	.533**	.713**	.674**	.674**	.702**	.194**	.032	.075**	.033	.085**	1.000	.221**	.172**	-.048	.131**	.409**
Debt/ Equity	.077**	.095**	.072**	.098**	.075**	.053*	.034	-.032	.088**	.042	.096**	.221**	1.000	.197**	.105**	.035	.124**
Debt/ Assets	.062*	.019	.076**	.086**	.071**	.056*	-.016	-.060*	.003	.078**	.022	.172**	.197**	1.000	.130**	.006	.103**
Earnings Vol	-.038	-.052	-.048	-.065*	-.045	-.076**	-.019	-.044	-.276**	-.074**	-.280**	-.048	.105**	.130**	1.000	.044	-.086**
Sensitive	.100**	.097**	.119**	.139**	.128**	.127**	.213**	.075**	-.053	-.133**	-.050	.131**	.035	.006	.044	1.000	.141**
Analyst Following	.354**	.335**	.468**	.551**	.411**	.484**	.366**	.153**	.256**	.021	.282**	.409**	.124**	.103**	-.086**	.141**	1.000

Table 5.8: Spearman Nonparametric Correlations with Outliers Removed

	Profit	Book Value	Market Cap	Revenue	Income	Equity	Foreign listing	US listing	Profit/Size	Earnings Margin	Return on Assets	Debt	Debt/Equity	Debt/Assets	Earnings Volatility	Sensitive	Analyst Following
Profit	1.000	.589**	.531**	.687**	.841**	.556**	.048	.037	.806**	.784**	.724**	.448**	.281**	.211**	-.311**	-.027	.561**
Book Value	.589**	1.000	.822**	.844**	.649**	.936**	.336**	.183**	.460**	.455**	.499**	.725**	.461**	.368**	-.256**	.094**	.742**
Market Cap	.531**	.822**	1.000	.695**	.569**	.812**	.336**	.193**	.427**	.384**	.461**	.521**	.269**	.185**	-.217**	.071**	.727**
Revenue	.687**	.844**	.695**	1.000	.793**	.740**	.160**	.089**	.574**	.517**	.652**	.710**	.493**	.408**	-.248**	-.024	.721**
Income	.841**	.649**	.569**	.793**	1.000	.595**	.097**	.038	.702**	.656**	.800**	.546**	.356**	.285**	-.289**	-.004	.637**
Equity	.556**	.936**	.812**	.740**	.595**	1.000	.358**	.203**	.432**	.434**	.449**	.603**	.389**	.219**	-.281**	.111**	.714**
Foreign listing	.048	.336**	.336**	.160**	.097**	.358**	1.000	.599**	-.018	-.027	.007	.170**	.041	.012	.026	.204**	.311**
US listing	.037	.183**	.193**	.089**	.038	.203**	.599**	1.000	.024	.007	.026	.037	-.033	-.058*	-.041	.075**	.170**
Profit/ Size	.806**	.460**	.427**	.574**	.702**	.432**	-.018	.024	1.000	.855**	.896**	.306**	.219**	.158**	-.265**	-.107**	.442**
Earnings Margin	.784**	.455**	.384**	.517**	.656**	.434**	-.027	.007	.855**	1.000	.740**	.336**	.250**	.200**	-.248**	-.137**	.391**
Return on Assets	.724**	.499**	.461**	.652**	.800**	.449**	.007	.026	.896**	.740**	1.000	.361**	.262**	.208**	-.272**	-.102**	.501**
Debt	.448**	.725**	.521**	.710**	.546**	.603**	.170**	.037	.306**	.336**	.361**	1.000	.805**	.839**	-.134**	.039	.541**
Debt/ Equity	.281**	.461**	.269**	.493**	.356**	.389**	.041	-.033	.219**	.250**	.262**	.805**	1.000	.817**	.020	-.017	.305**
Debt/ Assets	.211**	.368**	.185**	.408**	.285**	.219**	.012	-.058*	.158**	.200**	.208**	.839**	.817**	1.000	-.025	-.022	.243**
Earnings Vol	-.311**	-.256**	-.217**	-.248**	-.289**	-.281**	.026	-.041	-.265**	-.248**	-.272**	-.134**	.020	-.025	1.000	.044	-.223**
Sensitive	-.027	.094**	.071**	-.024	-.004	.111**	.204**	.075**	-.107**	-.137**	-.102**	.039	-.017	-.022	.044	1.000	.099**
Analyst Following	.561**	.742**	.727**	.721**	.637**	.714**	.311**	.170**	.442**	.391**	.501**	.541**	.305**	.243**	-.223**	.099**	1.000

5.3.8 Normality and Transformations

Examination of skewness and kurtosis values for most variables indicates that few if any of the variables follow a Normal distribution (table 5.1). The Kolmogorov-Smirnov and Shapiro-Wilk normality tests confirm this idea. For both tests, every variable tested has a significant deviation from a Normal distribution. While removing outliers tends to reduce the skewness and kurtosis the values are still very large (table 5.6) and this is reflected in the two tests still showing no Normally-distributed variables.

In order to limit the effects on non-normality on the testing, two transformations of the data are examined. Logarithmic transformation is used with the aim of reducing the non-normality of the sample while retaining potential relationships between variables. Where a variable can take a non-positive value, a constant is added before transformation to make even the lowest observed value positive so that logarithmic transformation is possible. When the transformed data are subjected to Normality tests, they are found to still be non-Normal. Visual inspection of plots of the variables reveals the problem to have two sources. Some variables are naturally skewed by having a minimum value of zero, such as the three measures of debt finance. Rather than a smooth bell-curve, such variables tend to demonstrate a high peak at zero due to many companies recording no debt. Logarithmic transformation of these zero values (after adding a constant of 1 to make the procedure possible) still leaves a large peak.

The other cause is in the addition of constants before transformation. Some of the constants are large enough that they dominate the new value that is then subjected to logarithmic transformation, leading all transformed values to be heavily influenced by the added constant. For example, if the smallest recorded profit is -£1m then the same amount is added to make all values positive. A profit (or loss) of lower (absolute) value than this has had a constant that may be several orders of magnitude larger than the original value added to it, meaning the £1m addition is more important in determining the logarithmically transformed value than the original observation. Transformed values therefore cluster around the logarithm of the added constant.

Overall, logarithmic transformation is of limited value for the data collected for this thesis. If variables were limited to positive values only then the addition of a constant would be unnecessary, reducing the problems this transformation causes. However, the use of potentially negative values in this sample (e.g. profit/loss values) means the choice is between leaving some non-normal variables untransformed or having the problems described above. Tables 5.9, 5.10, and 5.11 show details of the transformation. This transformation is applied to the full data set with outliers as the transformation should naturally result in more normally distributed variables.

Table 5.9: Descriptive Statistics with Logarithmic Data

	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Size measures						
Book Value (Log)	2.996	21.225	11.256	2.655	0.465	0.534
Market Cap (Log)	0.000	21.141	10.725	3.073	-0.778	2.770
Profit (Log)	6.306	19.868	11.253	2.093	1.084	0.625
Revenue (Log)	10.692	19.665	14.728	0.288	5.719	137.662
Income (Log)	0.000	18.997	13.725	0.615	-14.283	351.050
Equity (Log)	0.001	20.261	13.014	0.938	0.785	35.525
Foreign listing measure						
Foreign listing (# of countries)	0.00	6.00	0.81	0.78	.84	1.50
Debt finance measures						
Debt (Log)	0.000	19.901	6.879	5.120	-0.128	-1.189
Debt/ Equity (Log)	-11.148	5.144	-0.981	1.744	-1.765	4.391
Debt/ Assets (Log)	-11.614	3.479	-1.594	1.762	-1.555	3.223
Performance measures						
Profit/Size (Log)	4.361	5.006	4.700	0.013	-3.685	482.945
Earnings Margin (Log)	0.000	13.348	12.647	0.336	-37.292	1405.151
Return on Assets (Log)	3.256	3.689	3.609	0.019	-10.290	149.731
Volatility measure						
Volatility (Log)	-4.472	10.317	2.982	1.863	-0.335	0.423
Disclosure measure						
Following (# analyst estimates)	0.00	32.00	4.35	6.67	1.97	3.14

Table 5.10: Pearson Correlations with Logarithmic Data

	Profit	Book Value	Market Cap	Revenue	Income	Equity	Foreign listing	US listing	Profit/Size	Earnings Margin	Return on Assets	Debt	Debt/ Equity	Debt/ Assets	Earnings Volatility	Sensitive	Analyst Following
Profit	1.000	.234**	.095**	.054*	.142**	.100**	.102**	.081**	.177**	.044	.071**	.054*	-.014	-.037	-.016	.043	.169**
Book Value	.234**	1.000	.636**	.131**	.407**	.353**	.308**	.097**	.396**	.186**	.004	-.033	.014	-.045	.053*	.044	.403**
Market Cap	.095**	.636**	1.000	.055*	.223**	.177**	.155**	.045	.204**	.050	-.031	-.037	.015	-.058*	.044	0	.189**
Revenue	.054*	.131**	.055*	1.000	.279**	.256**	.324**	.628**	.383**	.172**	-.016	-.060*	.027	.019	.048	-.034	.306**
Income	.142**	.407**	.223**	.279**	1.000	.741**	.866**	.216**	.717**	.685**	-.002	-.199**	.155**	-.030	.328**	-.067*	.742**
Equity	.100**	.353**	.177**	.256**	.741**	1.000	.662**	.207**	.575**	.440**	-.061*	-.155**	.121**	-.027	.242**	-.027	.652**
Foreign listing	.102**	.308**	.155**	.324**	.866**	.662**	1.000	.273**	.694**	.682**	.029	-.173**	.109**	.032	.233**	-.082**	.784**
US listing	.081**	.097**	.045	.628**	.216**	.207**	.273**	1.000	.282**	.155**	.027	-.003	0.026	.015	.050	-.031	.275**
Profit/ Size	.177**	.396**	.204**	.383**	.717**	.575**	.694**	.282**	1.000	.474**	.005	-.100**	.051	-.031	.106**	-.041	.691**
Earnings Margin	.044	.186**	.050	.172**	.685**	.440**	.682**	.155**	.474**	1.000	.006	-.251**	.092**	-.021	.157**	-.048	.518**
Return on Assets	.071**	.004	-.031	-.016	-.002	-.061*	.029	.027	.005	.006	1.000	.911**	-.036	.020	-.073**	.050	.028
Debt	.054*	-.033	-.037	-.060*	-.199**	-.155**	-.173**	-.003	-.100**	-.251**	.911**	1.000	-.070**	.014	-.151**	.013	-.099**
Debt/ Equity	-.014	.014	.015	.027	.155**	.121**	.109**	.026	.051	.092**	-.036	-.070**	1.000	.005	.459**	.006	.073**
Debt/ Assets	-.037	-.045	-.058*	.019	-.030	-.027	0032	.015	-.031	-.021	.020	.014	.005	1.000	.005	-.022	.011
Earnings Vol	-.016	.053*	.044	.048	.328**	.242**	.233**	0.050	.106**	.157**	-.073**	-.151**	.459**	.005	1.000	-.028	.153**
Sensitive	.043	.044	.000	-.034	-.067*	-.027	-.082**	-.031	-.041	-.048	.050	.013	.006	-.022	-.028	1.000	-.055*
Analyst Following	.169**	.403**	.189**	.306**	.742**	.652**	.784**	.275**	.691**	.518**	.028	-.099**	.073**	.011	.153**	-.055*	1.000

Table 5.11: Spearman Nonparametric Correlations with Logarithmic Data

	Profit	Book Value	Market Cap	Revenue	Income	Equity	Foreign listing	US listing	Profit/Size	Earnings Margin	Return on Assets	Debt	Debt/Equity	Debt/Assets	Earnings Volatility	Sensitive	Analyst Following
Profit	1.000	.223**	.095**	-.020	.120**	.098**	-.006	.003	.134**	.058*	.077**	.060*	-.105**	-.136**	-.097**	.050	.120**
Book Value	.223**	1.000	.607**	.069**	.366**	.363**	.190**	.119**	.377**	.200**	.021	-.014	-.001	-.011	.029	.022	.340**
Market Cap	.095**	.607**	1.000	.049	.210**	.214**	.110**	.052	.224**	.064*	-.011	-.007	.029	.012	.036	-.033	.198**
Revenue	-.020	.069**	.049	1.000	.581**	.526**	.689**	.839**	.537**	.447**	-.101**	-.199**	.798**	.785**	.718**	-.305**	.553**
Income	.120**	.366**	.210**	.581**	1.000	.829**	.837**	.634**	.928**	.724**	-.081**	-.277**	.460**	.453**	.502**	-.247**	.739**
Equity	.098**	.363**	.214**	.526**	.829**	1.000	.698**	.559**	.812**	.526**	-.088**	-.195**	.426**	.387**	.463**	-.213**	.726**
Foreign listing	-.006	.190**	.110**	.689**	.837**	.698**	1.000	.791**	.726**	.711**	-.105**	-.343**	.579**	.533**	.658**	-.248**	.717**
US listing	.003	.119**	0.052	.839**	.634**	.559**	.791**	1.000	.570**	.539**	-.090**	-.227**	.698**	.658**	.795**	-.281**	.628**
Profit/ Size	.134**	.377**	.224**	.537**	.928**	.812**	.726**	.570**	1.000	.594**	-.191**	-.276**	.422**	.424**	.443**	-.278**	.707**
Earnings Margin	.058*	.200**	.064*	.447**	.724**	.526**	.711**	.539**	.594**	1.000	-.069**	-.414**	.307**	.336**	.357**	-.133**	.541**
Return on Assets	.077**	.021	-.011	-.101**	-.081**	-.088**	-.105**	-.090**	-.191**	-.069**	1.000	.819**	-.168**	-.109**	-.152**	.151**	-.083**
Debt	.060*	-.014	-.007	-.199**	-.277**	-.195**	-.343**	-.227**	-.276**	-.414**	.819**	1.000	-.198**	-.175**	-.210**	.100**	-.211**
Debt/ Equity	-.105**	-.001	.029	.798**	.460**	.426**	.579**	.698**	.422**	.307**	-.168**	-.198**	1.000	.859**	.897**	-.262**	.437**
Debt/ Assets	-.136**	-.011	.012	.785**	.453**	.387**	.533**	.658**	.424**	.336**	-.109**	-.175**	.859**	1.000	.744**	-.255**	.391**
Earnings Vol	-.097**	.029	.036	.718**	.502**	.463**	.658**	.795**	.443**	.357**	-.152**	-.210**	.897**	.744**	1.000	-.272**	.499**
Sensitive	.050	.022	-.033	-.305**	-.247**	-.213**	-.248**	-.281**	-.278**	-.133**	.151**	.100**	-.262**	-.255**	-.272**	1.000	-.213**
Analyst Following	.120**	.340**	.198**	.553**	.739**	.726**	.717**	.628**	.707**	.541**	-.083**	-.211**	.437**	.391**	.499**	-.213**	1.000

The other transformation tested is a Normal Score approach, as Cooke (1992) describes. The Normal score approach forces all variables to take a Normal distribution, but it eliminates the details of possible relationships among variables in the process. The first stage of this transformation requires observations to be ranked, removing the ability to examine how far two variables tend to move together and reducing possible relationships only to whether both increase in rank together or if one increases while the other decreases. The method then converts the ranks into values that force the variable to take a Normal distribution.

The Normal Score transformation data are not subjected to Normality testing. This data should generate a Normal distribution and testing was performed only in order to check the method had been performed correctly. For some variables the tests indicate non-Normality. On further inspection, the cause is equal values before transformation. Where two or more observations are equal they are each assigned the mean rank for all observations sharing the same value when the transformation ranks observations. In some cases, there are enough observations of a given value to distort the final Normal Scores. For example, a number of companies record zero debt in their capital structure, meaning that all of them share a rank for their debt, debt/assets, and debt/equity values.

One more transformation was investigated but not used, that of converting observations to their ranks. This was found to have the same problem discussed above with the Normal Score approach, eliminating information about the scale of relationships between variables as the only information retained by the transformation is the rankings and not the actual differences in size between any two observations. Further, this approach gives variables a new distribution that is effectively a discrete variation on the Uniform distribution, which does not help with later tests that assume a Normal distribution of variables. Finally, this approach tends to result in the same conclusions as the Normal Score method as both rely on ranking observations to some degree.

Tables 5.12 and 5.13 show the descriptive statistics and Pearson correlations for data after normal score transformation. The normal score transformation requires observations be ranked, so the resulting Pearson and Spearman correlations are nearly identical and only Pearson correlations are presented. This transformation is performed on the full data set for the same reason; the ranking stage of the transformation eliminates outliers by its nature, preventing them from causing problems.

Table 5.12: Descriptive Statistics with Normal Score Data

	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Size measures						
Book Value (Normal Score)	-3.391	3.391	0.000	1.000	0.000	-0.016
Market Cap (Normal Score)	-3.391	3.391	0.000	1.000	0.000	-0.016
Profit (Normal Score)	-2.180	3.391	0.003	0.992	0.059	-0.173
Revenue (Normal Score)	-3.391	3.391	0.009	0.979	0.121	-0.200
Income (Normal Score)	-3.391	3.391	0.000	1.000	0.000	-0.016
Equity (Normal Score)	-3.391	3.391	0.000	1.000	0.000	-0.016
Foreign listing measure						
Foreign listing (# of countries)	0.00	6.00	0.81	0.78	.84	1.50
Debt finance measures						
Debt (Normal Score)	-1.065	3.391	0.035	0.920	0.452	-0.471
Debt/ Equity (Normal Score)	-3.391	3.391	0.016	0.967	0.113	0.088
Debt/ Assets (Normal Score)	-1.065	3.391	0.035	0.920	0.452	-0.471
Performance measures						
Profit/Size (Normal Score)	-3.391	3.391	0.000	1.000	0.000	-0.016
Earnings Margin (Normal Score)	-3.391	3.391	0.000	1.000	0.000	-0.016
Return on Assets (Normal Score)	-3.391	3.391	0.000	1.000	0.000	-0.016
Volatility measure						
Volatility (Normal Score)	-2.009	3.391	0.004	0.988	0.086	-0.224
Disclosure measure						
Following (# analyst estimates)	0.00	32.00	4.35	6.67	1.97	3.14

Table 5.13: Pearson Correlations with Normal Score Data

	Profit	Book Value	Market Cap	Revenue	Income	Equity	Foreign listing	US listing	Profit/Size	Earnings Margin	Return on Assets	Debt	Debt/ Equity	Debt/ Assets	Earnings Volatility	Sensitive	Analyst Following
Profit	1.000	.234**	.095**	-0005	.126**	.108**	-0003	0023	.139**	.084**	0022	-0026	-.089**	-.151**	-.084**	0049	.169**
Book Value	.234**	1.000	.636**	.069**	.392**	.383**	.222**	.135**	.391**	.249**	.077**	.005	.013	-.015	.039	.015	.403**
Market Cap	.095**	.636**	1.000	.034	.219**	.210**	.119**	.039	.225**	.086**	-.008	-.054*	0035	-.006	.044	-.027	.189**
Rev.	-.005	.069**	.034	1.000	.506**	.480**	.626**	.794**	.464**	.428**	.226**	.147**	.690**	.683**	.610**	-.253**	.540**
Income	.126**	.392**	.219**	.506**	1.000	.817**	.824**	.563**	.883**	.738**	.444**	.277**	.434**	.393**	.481**	-.211**	.705**
Equity	.108**	.383**	.210**	.480**	.817**	1.000	.693**	.517**	.780**	.561**	.284**	.130**	.394**	.324**	.431**	-.173**	.710**
Foreign listing	-.003	.222**	.119**	.626**	.824**	.693**	1.000	.738**	.693**	.736**	.451**	.348**	.519**	.473**	.591**	-.211**	.694**
US listing	.023	.135**	.039	.794**	.563**	.517**	.738**	1.000	.506**	.532**	.289**	.212**	.607**	.561**	.688**	-.229**	.626**
Profit/ Size	.139**	.391**	.225**	.464**	.883**	.780**	.693**	.506**	1.000	.590**	.443**	.086**	.379**	.353**	.397**	-.226**	.665**
Earnings Margin	.084**	.249**	.086**	.428**	.738**	.561**	.736**	.532**	.590**	1.000	.687**	.748**	.264**	.306**	.307**	-.116**	.603**
Return on Assets	.022	.077**	-.008	.226**	.444**	.284**	.451**	.289**	.443**	.687**	1.000	.583**	.182**	.221**	.219**	.051	.276**
Debt	-.026	.005	-.054*	.147**	.277**	.130**	.348**	.212**	.086**	.748**	.583**	1.000	.083**	.163**	.105**	-.023	.191**
Debt/ Equity	-.089**	.013	.035	.690**	.434**	.394**	.519**	.607**	.379**	.264**	.182**	.083**	1.000	.774**	.892**	-.203**	.333**
Debt/ Assets	-.151**	-.015	-.006	.683**	.393**	.324**	.473**	.561**	.353**	.306**	.221**	.163**	.774**	1.000	.652**	-.208**	.290**
Earnings Vol	-.084**	.039	.044	.610**	.481**	.431**	.591**	.688**	.397**	.307**	.219**	.105**	.892**	.652**	1.000	-.224**	.371**
Sensitive	.049	.015	-.027	-.253**	-.211**	-.173**	-.211**	-.229**	-.226**	-.116**	.051	-.023	-.203**	-.208**	-.224**	1.000	-.139**
Analyst Following	.169**	.403**	.189**	.540**	.705**	.710**	.694**	.626**	.665**	.603**	.276**	.191**	.333**	.290**	.371**	-.139**	1.000

5.4: Methods

Studies of the determinants of disclosure among companies tend to use regression analysis as the primary means of statistical analysis. The discussion above in section 5.2 suggests this may not be the most appropriate method.

5.4.1 Regression and SEM Discussion

OLS regression is the technique commonly taught to students as the basic method of regression.

The method is easily explained using a case of simple regression where a dependent variable is explained in terms of a single explanatory variable. In such cases, plotting the regression on a 2-dimensional plane is possible and serves to clarify the process. In this case, the researcher may plot each observation to visually show the relationship between the two variables. Once each point is in place, a line of best fit is commonly drawn. However, it is very difficult to find this line precisely.

The regression process tests possible lines of best fit by identifying the distance between each point and the line being tested. In OLS, these distances are squared and summed. The line with the lowest total sum of squares is the final line of best fit. The squaring occurs in part to ensure that all values are positive before the summation, as using the original measured values would lead to a combination of positive and negative that would make most total distances near to zero regardless of whether the line passes perfectly through all points or is equally far above some points as it is below others. In addition, squaring values means that points far from the line contribute disproportionately highly to the total distance figure. The algorithm therefore favours lines that come close to all points over those that pass near-perfectly through most but far from a few.

While useful for ensuring that all observations are accounted for when minimising the error distance, this property can cause problems when a potential outlier is included in the sample. The line of best fit will be weighted towards this single point due to its large distance from any line that passes through all others. When squared, this distance becomes a very large discrepancy that the process needs to minimise.

The same approach is applicable to multivariate regression (i.e. multiple explanatory variables), albeit harder to visualise. The process effectively works the same way, however. A now hypothetical line is drawn between all observed points, as plotted in n-dimensional space, and the distance between the predicted point (the line) and the observed point (the plotted point) is calculated for each observation. These distances are squared and the reported regression equation is the one that describes the line with the lowest sum of squares.

Table 5.2 and 5.3 show correlations between size and multiple listing (Pearson ρ=0.175 with book value of assets), and between sensitive industries and multiple listing (Pearson ρ=0.234 with total foreign listing). While neither is particularly strong, there is an assumption in regression that all explanatory variables are completely independent of each other. The size-listing correlation is particularly important given the frequency with which these two variables are used together in regressions explaining disclosure. As Haavelmo (1943) argues, variables are often endogenous but researchers rarely consider the implications of this. As a consequence of the independence assumption, regression cannot easily be used on models in which explanatory variables have causal effects on each other (Iacobucci, 2009).

Violation of the independence assumption leads to some degree of inaccuracy in regression. Regression identifies the effect that each explanatory variable has on the dependent. If two explanatory variables have effects on each other, the relationship of each with the dependent is not clear. To use one of the examples above, assume the correlation between size and listing is because a large company has the resources to expand into foreign markets; size causes multiple listing. If both of these variables have effects on disclosure, as the literature suggests, then size has an indirect effect not accounted for in regression as it influences disclosure through its effect on multiple listing. Further, the multiple listing coefficient captures some of the effect of size. The coefficients contain some degree of inaccuracy, and there is a higher possibility of drawing erroneous conclusions as a result.

To get around the potential problems of correlated explanatory variables, alternative methods can be employed. Structural equations modelling (SEM) has been selected.

SEM is a generalised case of regression, using many of the same ideas to identify how a number of explanatory variables determine the values of other, dependent variables.

A single starting point for SEM is difficult to identify as it developed over time across multiple areas of research. Wright (1921) is one of the earliest known papers to focus on ideas that would later inform SEM. Essentially, Wright notes that there are cases where explanatory variables are statistically related to one another and may even have causal relationships. One of the examples used is a model predicting the weight of a guinea pig approximately one month after birth, in which many of the relevant variables are themselves influenced by the condition of the mother. A method is discussed involving correlations being recognised in the form of paths between variables, and is later applied to an example. The method bears similarity to the calculations performed in SEM and the same principles are in use, but the full methodology is yet to be developed at this stage. While potentially useful, the method described is computationally intensive. This was likely the reason for the paper's initially low impact; Wright's paper was published in 1921, decades before technological developments would result in practical computers. Despite this, Wright was able to make estimates in this and later papers.

Goldberger (1972) offers a useful overview of the development of SEM. Essentially, SEM has two models, one a structural model and the other a measurement model. These concepts were developed independently and brought together to form the method as it is now understood. Wright's (1921) path analysis concept provides the structural aspect, linking variables with expected causal links. At approximately the same time, psychologists began working with the method of Factor Analysis due to the recognition of unobservable variables underlying observed results. As a simple example, a subject's score on a series of tests may be explained by some unobservable mental talents, while a given disorder may be difficult to measure directly but easily observed through behaviours.

One of the largest steps in SEM was not the method itself being defined, but a computer program being developed for it. Joreskog created LISREL – short for LInear Structural RELations – during the 1970s. With this software, the computation-heavy nature of

SEM became less of a barrier to many researchers and the development of the method beyond its basic ideas received a major boost.

Despite overall enthusiasm, there are those who argue for simpler methods. Klein (1960) is a good example of this. The paper advises using simpler methods where possible and not becoming caught in the overenthusiasm that has allegedly followed since Haavelmo's (1943) work was published. In many cases, Klein argues, a more basic method will perform adequately without adding unnecessary complexity. More important, however, is to not regard complexity as a solution to all problems, which many were apparently doing at the time.

There are two primary benefits to SEM. First, unlike regression, there can be more than one dependent variable, and it is possible for any given variable to be both explanatory and dependent, being determined in part by other variables in the model (i.e. endogenous) while in turn serving as a determinant of others. Correlation of explanatory variables is therefore allowable and non-problematic in SEM. In addition to this, SEM can make use of latent variables. These are variables that cannot be observed directly, often due to a conceptual nature, but can be seen indirectly through a number of indicator variables. As an example, company size could be considered a latent variable. A company reliant on employee talent over equipment will have a comparatively small book value due to few recordable assets despite a large volume of trade and resulting high income and profit, likely leading to a large market capitalisation. At the other extreme, a heavily automated manufacturer will have a large total book value that largely defines the total market capitalisation, while the other common measures may be lower in comparison. More generally, various authors argue that multiple-indicator latent constructs are useful because individual measures will have some amount of bias or inaccuracy, but using several together should reduce the effect of any individual indicator's problems (Aaker and Bagozzi 1979, Diamantopolous 1994, Steenkamp and Baumgartner 2000). Cooke (1992) explains the use of multiple measurements for a concept in the disclosure context:

"There is no overwhelming theoretical reason to prefer one size variable to another. However, each variable may contain an interesting and possibly unique aspect of size, despite any multi-collinearity between the size variables."

Cooke, 1992, p232

Cooke (1992) uses factor analysis to resolve this problem, condensing several measures of company size into a single factor.

Similar cases can be handled in SEM by treating the various measures as indicators of a latent variable for the underlying concept. Using the same example, by taking a range of indicators, a more rounded view of the size of a given company can be obtained. The same holds for other variables commonly used in disclosure studies. For this reason, latent variables are used where possible for any variable where multiple measures were taken, although in many cases they proved unhelpful or problematic and are removed before the final model.

Figure 5.1 below shows a latent variable as it would appear in a SEM diagram. Arrows point from the indicator variables to the latent variable because the latent concept causes the values of its indicators. Further, the shapes given to the variables vary. Clarity in diagrams requires that latent variables be visually distinct from indicator variables. Established convention in SEM is for unobserved (i.e. latent) variables to take an oval shape while observed variables are rectangular.

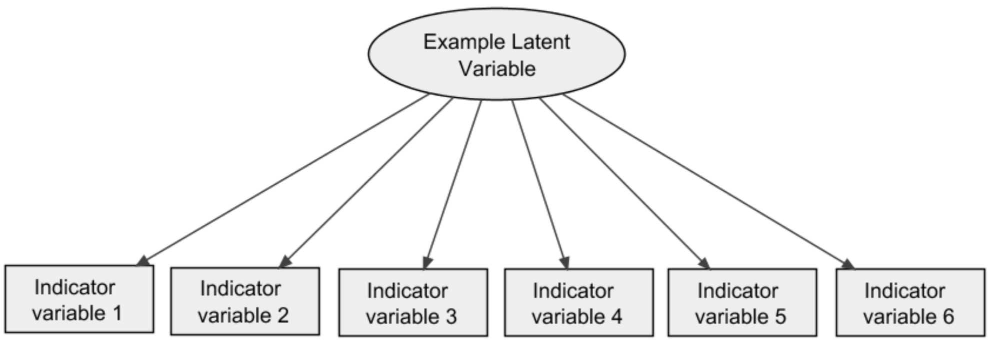

Figure 5.1: Latent variable example

By definition all indicators of a single latent variable should be correlated. This gives another means of including correlated variables in SEM, in addition to the ability to connect (latent) variables.

A second type of latent variable may also be used (Bollen and Lennox 1991, Jarvis et al 2003, Mackenzie et al 2005). The type above is known as a reflective latent variable as the indicators reflect the underlying concept. The other type is known as a formative latent as the indicators collectively define the underlying concept. The two types have very different properties and interpretations. None of the concepts used here were found to be better represented as a formative variable as all were either single-indicator variables (i.e. not latent variables) or underlying concepts better modelled with reflective latent variables. Appendix A discusses formative latent variables in more detail, along with additional SEM capabilities that were not considered necessary for this thesis.

Ultimately, the SEM process fits two models to the data. One examines how well the latent variables are measured by their indicators, known in some cases as the measurement model, and applies some confirmatory factor analysis to determine how strongly each indicator loads on the latent construct. Following this, the other tests the relationships between the latent constructs, known as the path model, which involves regression-like relationship tests (Aaker and Bagozzi 1979, Anderson and Gerbing 1982, Iacobucci 2009). As a result, the source of poor fit becomes comparatively easy to identify. While regression has the convenient measurement R^2 to indicate fit (see below), all sources of poor fit are encapsulated within $1-R^2$. In SEM, the two model approach makes clear whether poor fit is due to inaccurate measurement of constructs or inaccurate paths between constructs (Iacobucci, 2009).

Figure 5.2 below presents an example of a structural equation model.

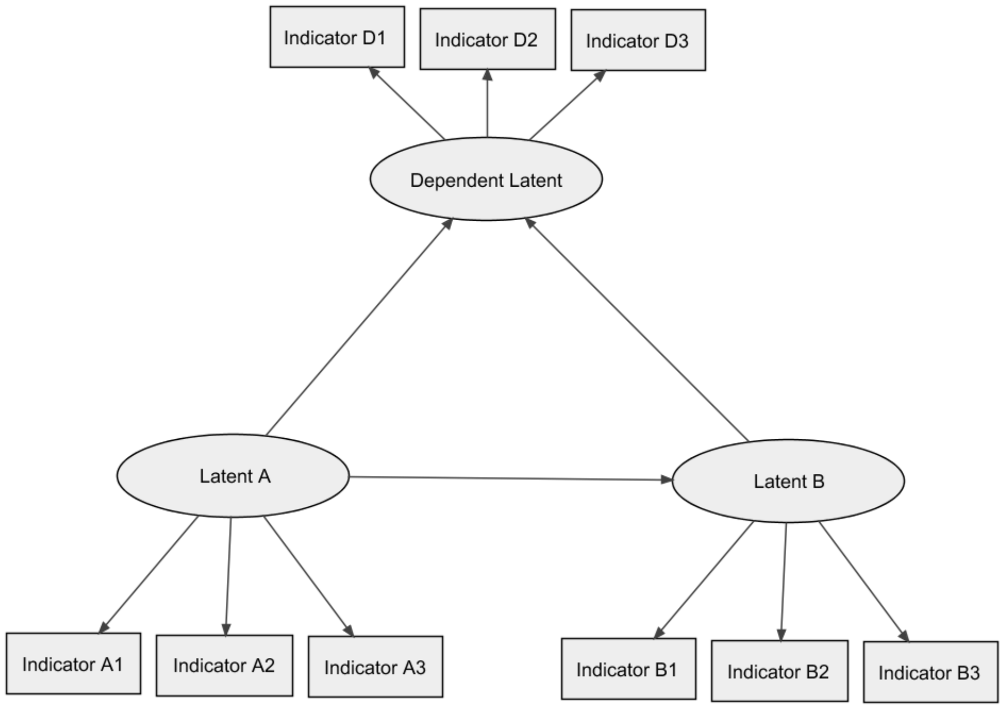

Figure 5.2: Example SEM

This diagram shows a number of conventions of SEM. As described above, observed variables (the indicators) are represented with angular boxes, while latent variables are elliptical. Arrows represent causal links, so in this case Latent A is thought to influence both Latent B and the dependent variable. There is no requirement that only latent variables be used; if one of the latent variables can be adequately represented with a single indicator then it can be replaced with that indicator.

Finally, it should be noted that this diagram (like all the model diagrams in this thesis) is, strictly, incomplete. Any endogenous variable, defined as one caused to any extent by another variable present in the model, should have another causal variable added to it. This represents an error term, i.e. the part of its value that is not explained by causes present in the model. In Figure 5.2, only Latent A would not have such a variable attached as it is the only exogenous variable in this model; the other two latent variables and all of the indicators would have an error. These error terms are usually represented with a circular shape to distinguish them from other types of variable. Throughout the thesis, these error terms are left out of model diagrams as an aid to clarity. They were included during the analysis of each model.

There are some disadvantages to SEM, however. First, it is complicated; regression is simpler and may be a better option despite its problems if they are known and worked around. In cases where the regression assumptions are minimally violated, regression is a less complicated method that gives reliable results. It is possible that the regression models are accurate enough that the effort necessary for SEM is unnecessary.

Second, there is not a single convenient measure of model fit in SEM, unlike regression which has R^2 as a clear indicator of fit. Instead, the normal recommendation is to use a series of fit measures for a comprehensive overview of overall fit. In addition to this, the chi-square test – one of the more useful fit measures – is sensitive to sample size. With larger samples it becomes more likely to indicate a poor model fit despite large samples being otherwise preferable in statistical studies.

Due to the question of the appropriateness of SEM and the need to compare it to existing techniques, it is not the only modelling process used in this thesis. As seen in Chapter 6, regression models of the type commonly seen in disclosure literature are tested first as a basis for comparison. Chapter 7 covers most of the results of SEM approaches to modelling and contains a comparison of the results of the two methods, although there are a few exploratory structural equation models in chapter 6.

The methods used, both SEM and OLS, are forms of estimation. That is, they attempt to find a value for some unobserved parameters based on the observed data. In each case, the main items estimated are the extent to which some variables influence the value of others.

In regressions, the only estimation required is the coefficient attached to each of the explanatory variables, the various β_i in the regression equations. For SEM, estimation is more complex. All arrows in a path diagram, both straight causal links and curved covariances, are estimated.

A straight connection between two non-indicator variables is conceptually similar to a β_i coefficient in a regression in that it estimates the extent to which the value of one variable influences the value of another. Straight arrows between a latent variable and its indicators are again similar as they estimate the extent to which the latent variable's unknown value influences that of the indicators. However, for each latent variable, the estimated weight of one indicator will be constrained to be exactly 1 in order to provide a scale for other weights. This does mean weights are defined in relation to each other rather than on a more absolute scale and the values will change if a different indicator is used for scaling.

Curved covariance arrows are conceptually different to the straight arrows more commonly seen in structural equation models. Where a curved arrow is not present between two variables, the researcher is assuming zero covariance. The existence of an arrow relaxes this assumption, but it is possible for the calculated covariance to be close to zero or statistically insignificant.

SEM estimation is a more sophisticated and data-intensive method than regression analysis. At its core, the method relies on covariance matrices. Parameter values for the model are estimated using the covariance matrix Σ containing implied values of covariance obtained from the starting values for required parameters (themselves obtained by differential calculus). This matrix is compared against the covariance matrix S, which contains the observed values. A maximum likelihood approach is then employed to find a sufficiently probable Σ given the observed S, which is then used to determine the parameters required to generate the final matrix Σ.

The actual maximum likelihood fitting function is:

$$F_{ML} = \log|\Sigma| + tr(S\Sigma^{-1}) - \log|S| - (p + q) \quad (1)$$

Where tr indicates the trace of the matrix $S\Sigma^{-1}$ and $(p + q)$ is the total number of variables in the model. The maximum likelihood is obtained where this function is minimised. Bollen (1989) explains the derivation of this equation.

For OLS regression, the estimation process follows a simple formula. The error term ε_i in the regression equation should be minimised for all i. By rearranging the regression formula for ε_i, the estimation minimises:

$$\varepsilon_i = Y - \alpha - \beta_1 x_1 - \beta_2 x_2 - \ldots - \beta_n x_n \quad (2)$$

A number of estimation techniques are possible in SEM. The chosen method here is maximum likelihood estimation (MLE). This approach to estimation uses the observed data to find the most likely (i.e. highest probability) values for statistical parameters that describe the observed sample's distribution.

5.4.2 Possible Alternative Methods

The main reason for the use of an alternative modelling technique is a perceived endogeneity problem in the data commonly used in regression studies that explain disclosure activity. The chosen SEM approach is but one of many possible techniques that allows for some degree of endogeneity in models.

One of the most obvious alternatives is an instrumental variables (IV) regression approach (and the related 2-stage least squares method). IV regression includes at least one instrumental variable in analysis. In a simple regression of the form $Y = \alpha + \Sigma(\beta_i X_i) + \varepsilon$, an instrument for X would be independent of Y but not X. The value of the instrument does not influence Y directly, but it does influence X which in turn affects Y. If X can be held constant while changing the instrument's value, Y should also remain constant. An instrument for explanatory variable X can be loosely defined as one which influences the dependent Y only through effects it has on X.

IV analysis is useful in cases where the value of an explanatory variable X is correlated with the error term ε. This often means there is some correlation among explanatory variables. While this describes the situation with the explanatory variables commonly seen in disclosure studies, IV analysis is not used here. To some extent this is simply because of the benefits of SEM discussed above (mostly in latent variables). Beyond this, using IV analysis requires being able to find an instrument, which may be a challenge. The properties required of an instrument are such that it can be very difficult to find a suitable variable to fill this role. While it may be possible to find a variable that would serve as an instrument for company size, this would not be easy and would not enable the use of latent variables.

In terms of the underlying mathematics, the models that rebuild regressions in SEM are applying regression techniques. The SEM process fits two models sequentially. One, the measurement model, tests the latent variables to determine how well they are described by their indicator variables. The second model, the structural model, then tests relationships between non-indicator variables using regression techniques. These can involve latent variables only, latent and observed (non-indicator) variables, or observed variables only. In the case of a model with only observed variables, there is no measurement model to test and the process moves straight to the structural model.

The structural model phase is capable of handling a range of complex relationships between variables. However, in the case of the remade regression models in this thesis, there are no complex relationships. Six observed variables are assigned as causes of a single observed dependent and nothing else is included in the model. All that needs to be calculated is the extent to which each variable explains the dependent, exactly as in a simple OLS regression. Differences in values between one regression and its SEM remake are generally small and likely explained by rounding errors somewhere in the process as the two programs may use a slightly different algorithm or work to more significant figures in the process even if they display to the same number.

In addition to different methods, different estimators could be used with SEM such as Generalised Method of Moments (GMM). This covers a family of estimators with the benefit that some estimator can be found within the GMM framework that should provide a good estimation for any data set.

MLE, the default in many programs that cover such techniques, is an efficient and unbiased estimator as long as the data meets certain requirements, primarily in terms of meeting assumptions about an underlying normal distribution. Where the data does not meet these requirements, GMM techniques will provide a better estimator. MLE is also computationally less complex and therefore faster to calculate, which was more advantageous before computing power became widespread and easily accessible. Both of the SEM programs used (Stata 14 and AMOS 22) include the option to use asymptotic distribution-free (ADF) estimators, which are part of the GMM category.

Investigation of the dependent variable above shows that it does not deviate far from a normal distribution. In the regression models in chapter 6, its errors (residuals) are demonstrated to similarly not deviate far from a normal distribution. MLE is known to be highly efficient under these circumstances and is therefore the preferred form of estimation throughout.

A latent variable in SEM is used where there is a hypothetical variable that cannot be measured directly but would be useful for the model under consideration. A latent variable is formed by identifying a number of indicators, measurable variables that will have their values influenced by the unseen value of the latent variable. In practice, this means that a number of observable variables are condensed into a single unobserved one in the structural model phase. There is a clear comparison to be made with factor analysis.

Factor analysis is normally used to identify whether a group of variables can be reduced to a smaller number of unobserved variables that each represent the commonalities between subsets of the observed variables. Its purpose is to reduce a large, complicated set of variables down to a smaller, more manageable number. There is also a variation on the technique known as confirmatory factor analysis. This is used where a researcher has a pre-existing construct from theory or prior examination and wishes to test that their expected measurement variables are the correct set (e.g. identifying whether measures should be added or removed). A SEM latent variable can be examined in this way; the researcher will usually have pre-existing ideas about where a latent variable exists and can use CFA to test these ideas. This process is performed as part of the measurement model stage of SEM.

As an additional test here, an exploratory factor analysis was performed. All of the observed variables were used in a single test to investigate whether the expected latent variables emerge from this method. Overall, the EFA is consistent with the *a priori* expected latent variables. The factors that emerge generally match the expected latent variables in terms of the measures that show large loadings. The main difference is that there is some evidence that book value of assets and total debt together load to some extent on a previously unexpected factor. It is not immediately clear what this factor represents other than some aspect of company size. While this potential factor is not used in later modelling, the size latent variable often has relatively low loading from book value compared to other size measures, and the debt capital latent is usually well-represented by the debt/assets and debt/equity measures alone.

5.5: Conclusion

Data has been gathered for a range of variables. Examination of the data in section 5.3 shows some potential problems, although all are to be expected from the data. The variables are tested for correlations, which show some apparent links between variables in some cases. In general, high correlations occur within the variable groups defined by the headings above, indicating that different measurements of the same concept tend to be correlated as expected. Measurements of different concepts rarely have high correlations, suggesting few cases of one variable potentially acting on both. Importantly, many of the variables are correlated significantly (if not always highly) with the dependent variable, suggesting that most may have an effect on disclosure practices.

However, these are conclusions drawn from correlation analysis alone. The next chapter covers the results of regression techniques involving the data, while chapter 7 covers SEM approaches.

Chapter 6: Regression Modelling

The layout of this chapter is explained here. Section 6.1 discusses the regression process and models used, and the major assumptions made in this form of testing. Section 6.2 demonstrates and discusses the results of the regression analyses performed on the data. Sections 6.3, 6.4, and 6.5 repeat this process for the outlier-removed data, the logarithmic data, and the normal score data respectively. Section 6.6 discusses the overall findings taking all four sets of models into account.

Section 6.7 shows the initial results of SEM methods and includes rebuilding the regressions as simple structural equation models, then goes into detail on various steps taken with the structural models and discusses a single one in depth as an example. This process always results in near-identical coefficients as the OLS regressions. The results of this process are reported only for the basic data models as this is sufficient to illustrate the comparability.

As a general rule, statistical significance is accepted here at the 5% level. However, variables significant at 10% are kept in the model when insignificant items are deleted. This is essentially to allow for the possibility that further changes to the model may yet show significance to the item in question. No conclusions are drawn that assume such items are supported by the analysis, however. A few references to a “grey area” of significance are made, which means 0.05<p-value<0.10.

6.1: Regressions Basics and Assumptions

Research in this field frequently uses regression analysis to determine what combination of characteristics make a company likely to disclose information. A similar analysis is carried out here as a baseline for comparisons and to forge expectations for later models. In keeping with past research, all regressions performed are basic OLS regressions. Other techniques are available but rarely used in the area, with the exception of logit methods where the dependent is a binary variable and Tobit methods where there is some censoring of the dependent.

Various pieces of research use different variables to measure the same concept. In some cases, these alternatives are effectively mutually exclusive due to regression methods encountering distorted results where two or more included variables are strongly related to one another. Using company size as an example, some papers use the book value of assets as a measure of size (e.g. Singhvi and Desai 1971) while others recommend a measure based on the company's revenue (e.g. Adams et al 1998). Both are valid measures of size despite giving different information about the company and it is not a simple matter to identify a superior one. Each is ultimately a proxy measure and will therefore estimate the size of some companies more accurately than others.

As an example, companies are generally unable to value human capital on the balance sheet. In a company reliant on employee abilities as a major source of income, the book value of assets will be small relative to the revenue of the firm. On the opposite side, a heavily automated manufacturing company will have a high asset value simply because its primary 'workforce' consists of machines that can and should be considered assets, making the assets high relative to the revenue.

In order to deal with the fact that various measures are suggested for several variables, a number of regressions have been performed. This additionally serves as a robustness test, ensuring that observed effects are explained by what the variables represent instead of some statistical quirk of the data, for example showing that size is the cause instead of revenue specifically. There are three regression models in total, each using a different rule for variable selection where there are options. Regression 1 uses the most commonly employed measure of each variable, i.e. the measurement seen in the largest number of prior papers examined. Regression 2 uses the second most common measure in each case. Regression 3 deviates from the pattern. Where a third variable is available, this is used; otherwise, the first choice is re-used.

All regressions take the same basic form:

$$DISC = \alpha + \beta_1 SIZE + \beta_2 LIST + \beta_3 PERF + \beta_4 DF + \beta_5 SENS + \beta_6 VOL + \varepsilon \qquad (3)$$

Where:
DISC, the dependent variable, is the measure of disclosure. This is the analyst following measure described in chapter 5
SIZE is a measurement of the company's scale. The three measures used are book value of assets, market cap, and revenue.
LIST is a variable for the company's multinational listing status. There are two measures here. The first is a count of the total number of countries in which the company is listed. The other is a dummy variable for whether the firm is listed in the USA.
PERF is a performance ratio for the company. Three are employed: the profit/size ratio; return on assets; and the earnings margin.
DF measures the level of debt involved in the company's capital structure. Three measures are used: the Debt/Assets ratio, Debt/Equity ratio, and the monetary value of debt in the company.
SENS is a dummy variable set to 1 if a company is in a publically sensitive industry sector, as explained in chapter 5. This measure never varies between models as there is only one measurement of sensitivity.
VOL measures the company's earnings volatility. This is calculated as the standard deviation of return on equity over a five-year period (or less where data is unavailable for the full period, as explained in chapter 5). This measurement is unchanged between regressions.
The Greek letters are all standard regression notation. The α represents a constant of unknown value before analysis, the various β_i (i=1,2,…,6) are coefficients of unknown value before analysis, and the ε is an error term with a Normal(0,σ) distribution (σ being an unknown standard deviation) that reflects some random variation around the model's predictions.

In addition to the analysis detailed below, all models were checked for multicollinearity of variables using the variance inflation factor (VIF) measure. In all cases, the VIFs are below 2. Even the strictest of tests using this variable suggest it should not exceed 5. As a result, no further mention is made as all regression models are assumed to not have any problems with correlated variables.

Further, three major assumptions made in regression analysis are tested and discussed below. The first assumption is that variables are measured on a continuous scale. This is true for only a few variables in this study, with the only unambiguously continuous variables being those that are formed by calculation, such as the debt and performance ratio measures. US listing is the most notable violation of this assumption as it is a dummy variable taking values of 0 or 1. Models that do not use this variable instead include the total number of foreign listings, which takes discrete values ranging from 0 to 6. Analyst following takes a wider range, but still by its nature must take discrete values. Any variable that represents a currency quantity, e.g. any of the size measures, presents a less severe violation. These variables are generally measured in units of £1m and are rounded, meaning they are discretely measured. However, these variables take values in a wide enough range that they can be practically considered continuous.

The second assumption tested is that there is a linear relationship between the dependent variable and each dependent variable. While this can be investigated by plotting the two, this is rarely helpful in this thesis. When most variables are plotted against analyst following, the resulting diagram does not provide much information on the relationship. Many companies have a low following, so regardless of the other variable's value there is a general tendency for the plots to be largely flat.

Correlation analysis proves more helpful in this regard. Comparing tables 5.2 and 5.3 shows that, for most variables, the Spearman correlation with analyst following is much stronger than the same variable pair's Pearson correlation. This suggests that nonlinear relationships are generally far stronger than linear relationships. While linear relationships cannot be ruled out, it is likely that they generally do not accurately describe the relationship between the dependent and independent variables.

Logarithmic transformation does not change this result much, reducing the scale of the difference between Pearson and Spearman correlations while still showing stronger Spearman correlations. Normal scoring makes the differences much smaller, but still favours the Spearman correlations. The two should actually be identical, but the differences are minor and likely caused by the Spearman correlation algorithm handling identical values differently to the ranking that occurs in the Normal Score transformation.

The third assumption tested is that the residuals produced from the regressions follow a Normal distribution. This is quickly proven false by examination of a histogram of the residuals of each regression, as presented directly beneath each regression below. For the basic, untransformed data including possible outliers, the residuals consistently display a slight positive skewness and very definite leptokurtosis in each case. The skewness indicates a likely violation of another assumption, that of no significant outliers after the regression (i.e. no large deviations from the regression line). The higher than expected number of large (positive) residuals that the skewness indicates implies that some residual values may be far higher than expected and qualify as outliers.

Removing outliers from the sample reduces the kurtosis, but both it and the skewness are still clear after regression. Logarithmic transformation produces inconsistent results in this regard. The residuals of regression 1 are slightly leptokurtic but otherwise close to Normal. Regression 2 displays some skew and very clear leptokurtosis, while regression 3 is between the other two in that it shows minor skew and moderate leptokurtosis. The rank and normal scores transformations both have similar results in all models. These transformations result in minor positive skew and slight leptokurtosis.

The final assumption discussed is that variables are measured without error. This is not strictly a regression assumption as most statistical methods will benefit from reliable measurement. As discussed in section 5.3.7 when addressing outliers in the sample, the data here should be considered measured without error. The sample consists of listed companies and most observations are values that were recorded in the companies' annual reports. The reliability assumption emerges from the audit process that occurs before the annual reports are published. Auditors cannot prevent all errors; even excluding cases where fraudulent reporting has been able to occur for several years, the auditors cannot reasonably examine all figures in great detail, and some published figures rely on a degree of judgement from within the company (provisions for bad debts are a clear example of this; although not part of this study, the value of such provisions may influence other recorded values). However, the auditors' role is to prevent material misstatements. Values in annual reports may be in error to some extent, but should be free of errors of any notable scale.

6.2: Regressions on Base Data

6.2.1: Regression 1

The first regression model tested as an explanation of disclosure activity is:

$$DISC = \alpha + \beta_1(\text{Book value of assets}) + \beta_2(\text{Total foreign listing}) + \beta_3(\text{Return on assets}) + \beta_4(\text{Debt/Assets}) + \beta_5(\text{Sensitive industry}) + \beta_6(\text{Volatility}) + \varepsilon \quad (4)$$

The results of this regression are given in the table below. The R^2 fit indication is 0.230. Similar research tends to obtain values around 0.3-0.4 range (e.g. 0.306 for Raffournier (1995) and Jiang et al's (2011) 0.370). The obtained 0.230 is rather low in absolute terms, but only slightly lower than comparable research.

Table 6.1: Results of Regression 1

Model Data	F-stat	Model Sig.		R-square	Adj. R-square
	72.335	0.000		0.233	0.230
Variable Data	**Std. Coeff.**	**Coeff.**	**Std. Error**	**t-stat**	**Sig.**
Constant		1.442	0.255	5.606	0.000
Book Value	0.204	1.74E-08	0	8.677	0.000
Foreign listing	0.343	2.934	0.207	14.144	0.000
Sensitive	0.077	1.023	0.319	3.205	0.001
Return on Assets	0.161	1.700	0.261	6.523	0.000
Debt/ Assets	0.064	0.416	0.16	2.596	0.010
Volatility	-0.012	0.000	0.001	-0.513	0.608

The only coefficient to fail the significance test at the 5% level is that of volatility, which is consistent with its very small standardised coefficient of -0.012.

The coefficient of Book Value is similarly small, but this is an effect from the scale of the variable. In this case, the standardised coefficient provides more information; removing the effect of variable size demonstrates that this is actually one of the more powerful coefficients with a value of 0.204, beaten only by Foreign Listing's standardised coefficient of 0.343.

The remaining coefficients do not need much comment. Sensitive industry membership adds one analyst above the average for a company with similar characteristics in an insensitive line of business. The ROA coefficient appears large, indicating some analyst interest from performance. Finally, the use of debt generates analyst interest, but at a lower rate than other potential sources.

However, upon further analysis, it becomes clear – as mentioned earlier – that the regression is not working properly. The diagram below shows a histogram of the residual values.

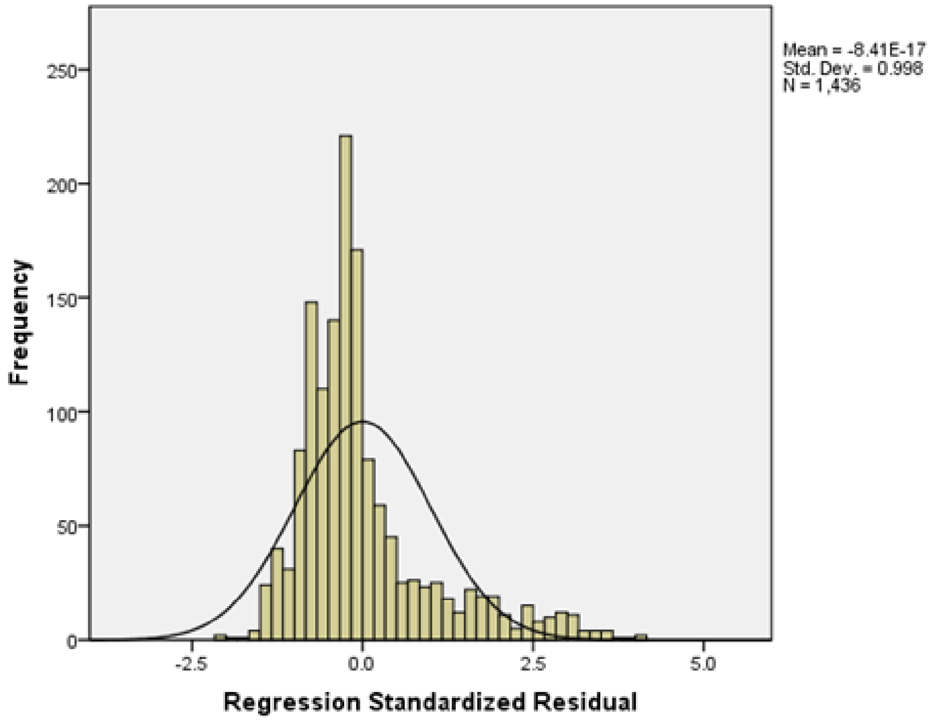

Figure 6.1: Residual Histogram of Regression 1

The line represents what a Normal distribution of the same mean and standard deviation as the residuals would look like. The histogram clearly deviates heavily from this. The peak is off-centre, indicating a nonzero skewness. In addition, the peak is obviously far higher than it should be. Combined with the long right-hand tail, this suggests some degree of leptokurtosis. This is not entirely unexpected; analysis of the variables individually generally resulted in high kurtosis and skewness.

6.2.2: Regression 2

This model uses less commonly used variables. The resulting equation is:

$$DISC = \alpha + \beta_1(\text{Revenue}) + \beta_2\text{US Listing}) + \beta_3(\text{Earnings Margin}) + \beta_4(\text{Debt/Equity}) + \beta_5(\text{Sensitive industry}) + \beta_6(\text{Volatility}) + \varepsilon \quad (5)$$

The regression results are displayed below. R^2 this time is 0.122, a clear indication of a poorly-fitting model.

Table 6.2: Results of Regression 2

Model Data	F-stat	Model Sig.		R-square	Adj. R-square
	34.220	0.000		0.126	0.122
Variable Data	**Std. Coeff.**	**Coeff.**	**Std. Error**	**t-stat**	**Sig.**
Constant		2.874	.238	12.068	.000
Revenue	.256	1.06E-7	.000	10.275	.000
US listing	.148	2.513	.424	5.922	.000
Sensitive	.140	1.867	.334	5.584	.000
Earnings Margin	.010	4.97E-6	.000	.407	.684
Debt/ Equity	.001	0.001	.030	.043	.966
Volatility	-.042	-0.002	.001	-1.672	.095

This arrangement of variables changes things compared to regression 1. Volatility remains insignificant (albeit in the 'grey area' where 0.05<p-value<0.10) but is joined by the debt finance and performance measures (debt/equity and earnings margin, respectively). The significant determinants of disclosure under this model are sensitive industry membership, US listing, and size as measured by reference to revenue.

However, residual analysis suggests severe problems as seen in Figure 6.2 below:

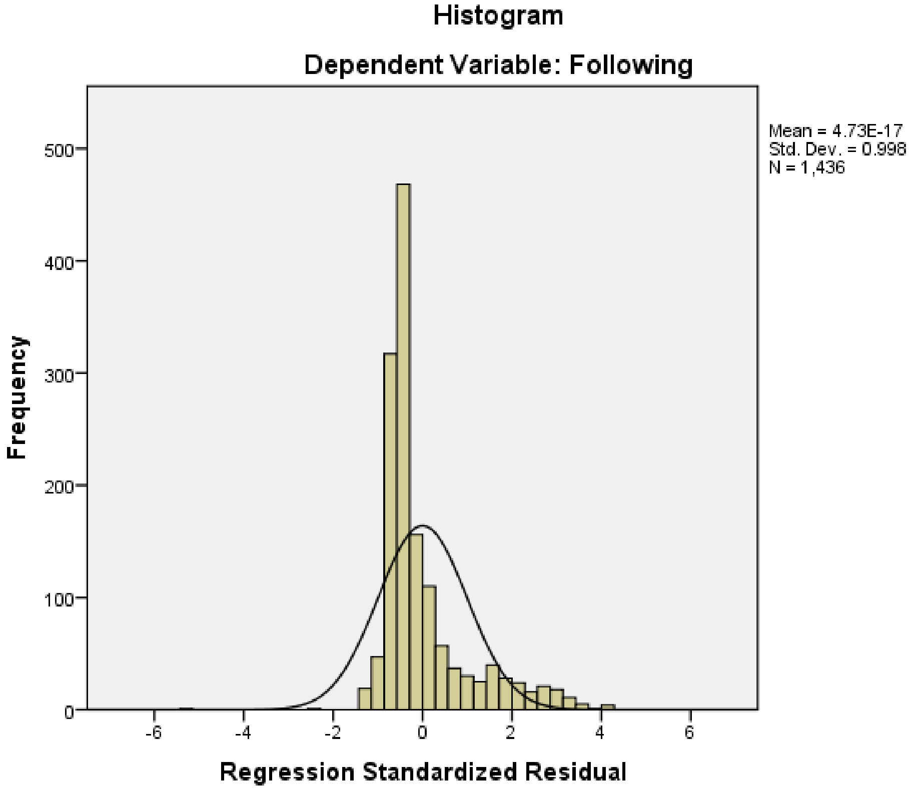

Figure 6.2: Residual Histogram of Regression 2

Compared to regression 1, these residuals are similarly skewed and much more leptokurtic. All evidence indicates that regression 1 is superior.

6.2.3: Regression 3

This third model continues the pattern set by the second, using the third most common measurement of a given variable where such could be found. The model is:

$$DISC = \alpha + \beta_1(\text{Market Cap}) + \beta_2(\text{Total foreign listing}) + \beta_3(\text{Profit/Size}) + \beta_4(\text{Debt}) + \beta_5(\text{Sensitive industry}) + \beta_6(\text{Volatility}) + \varepsilon \quad (6)$$

The listing measurement has been set as the total number of listings again, as in regression 1, because there was no third option. The debt finance variable this time is the monetary amount of debt in the firm. This variable is questionable; while it gives a clear indication of the amount of debt, it does not scale to the firm's size. Further, in testing correlations among variables, it became clear that this correlates more strongly with size measures than other debt finance indications. This gives some cause for concern about co-linearity problems, although no such problem emerged as the VIFs remained below 2.

The fit for this version of the model is an R^2 of 0.211, putting it in the middle of the three regression models.

Table 6.3: Results of Regression 3					
Model Data	F-stat	Model Sig.		R-square	Adj. R-square
	64.859	0.000		0.214	0.211
Variable Data	Std. Coeff.	Coeff.	Std. Error	t-stat	Sig.
Constant		1.405	.255	5.506	.000
Market Cap	.092	1.46E-8	.000	3.853	.000
Foreign listing	.350	2.997	.209	14.315	.000
Sensitive	.075	1.000	.323	3.093	.002
Profit/Size	.066	0.299	.106	2.810	.005
Debt	.168	6.94E-8	.000	7.010	.000
Volatility	-.028	-0.001	.001	-1.204	.229

As is becoming consistent, volatility is not significant at the 5% level. All other variables are. Market cap and debt both have very small coefficients, but again as monetary amounts they are measured on such scales that the net effect of multiplying the coefficients by the variable values will in many cases mean a small but noticeable effect. Otherwise, there is little to add that has not been applicable to other models so far.

As with the previous two models, analysis of residuals indicates some troubles with the regression method used here:

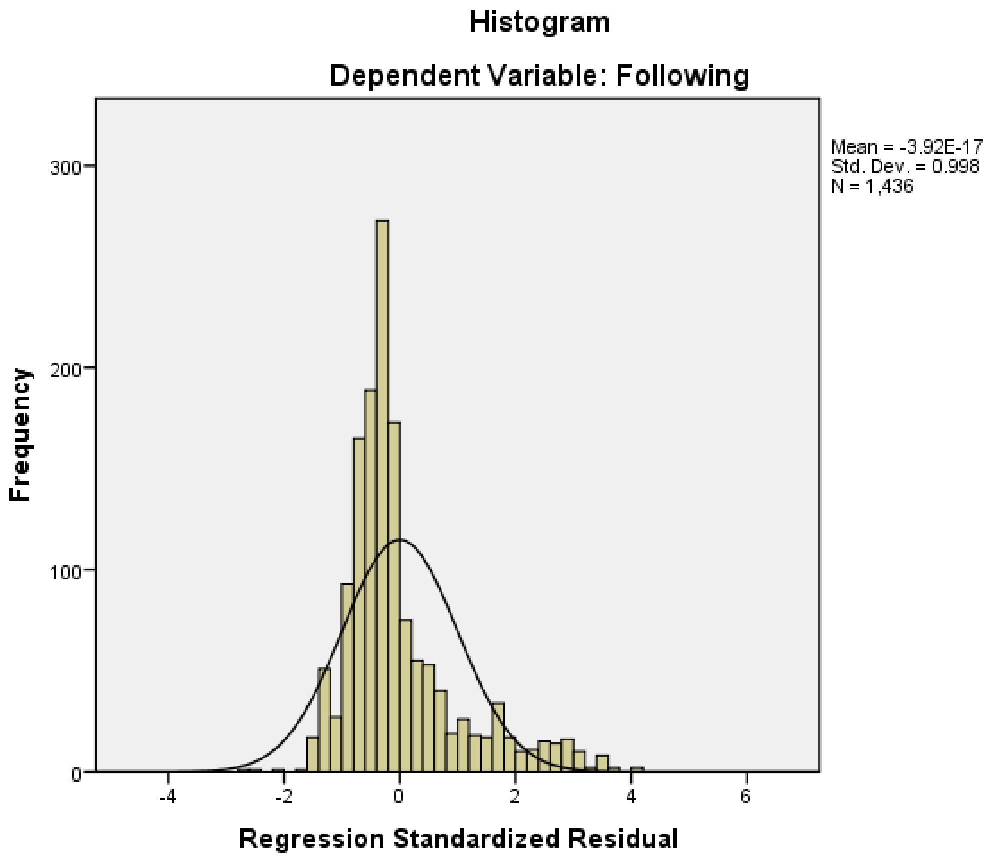

Figure 6.3: Residual Histogram of Regression 3

Once again, there is a clear positive skew and very definite leptokurtosis. The residuals appear similar to those of Regression 1 in Figure 6.1 above.

6.3: Regressions with Outliers Removed

6.3.1: Regression 1

Table 6.4: Results of Regression 1 with Outliers Removed					
Model Data	F-stat	Model Sig.		R-square	Adj. R-square
	88.617	0.000		0.281	0.277
Variable Data	Std. Coeff.	Coeff.	Std. Error	t-stat	Sig.
Constant		1.378	0.263	5.243	0.000
Book Value	0.246	0.000	0.000	10.400	0.000
Foreign listing	0.289	2.390	0.199	12.002	0.000
Sensitive	0.068	0.866	0.300	2.892	0.004
Return on Assets	0.255	6.110	0.576	10.610	0.000
Debt/ Assets	0.099	2.807	0.660	4.253	0.000
Volatility	-0.012	-0.001	0.003	-0.495	0.621

Retesting regression model 1 after removing the outliers from the sample improves the model fit by a small amount (R^2 0.277). In terms of the important variables, foreign listing still has the largest standardised coefficient, but not by as large a margin. In addition, return on assets has become more powerful than the company's book value. Overall, this model most supports a Signalling explanation of disclosure due to the importance it places on the financial performance variable, ROA. It does not reject either of the other theories, however.

Inspection of residuals below shows that, while this change to the data improves the resulting normality, it does not resolve the problems. The diagram shows the residuals to be less skewed and less leptokurtic than in the original version of regression 1, but still showing both of these traits.

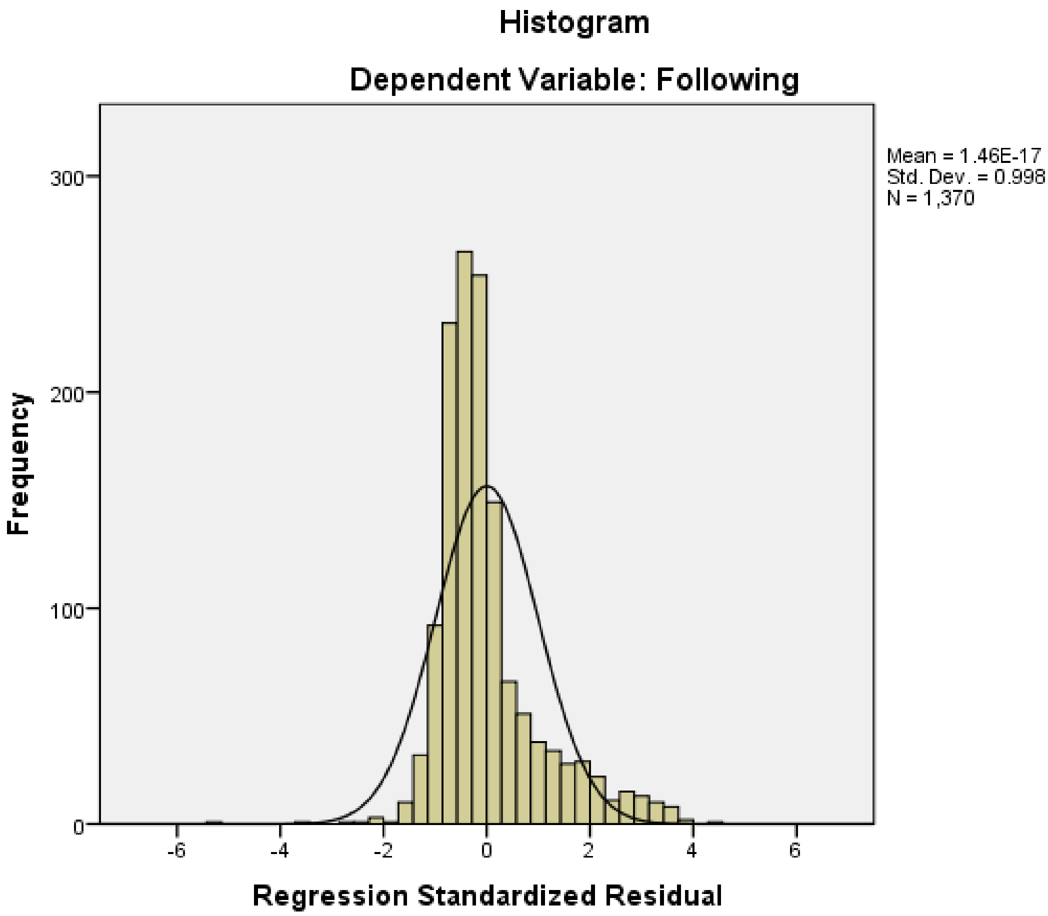

Figure 6.4: Residual Histogram of Regression 1 with Outliers Removed

6.3.2: Regression 2

Retesting regression model 2 with outliers removed shows some very different outcomes compared to the first version.

Table 6.5: Results of Regression 2 with Outliers Removed					
Model Data	F-stat	Model Sig.		R-square	Adj. R-square
	111.217	0.000		0.329	0.326
Variable Data	Std. Coeff.	Coeff.	Std. Error	t-stat	Sig.
Constant		2.649	0.213	12.412	0.000
Revenue	0.522	0.000	0.000	23.022	0.000
US listing	0.114	1.848	0.365	5.057	0.000
Sensitive	0.061	0.776	0.288	2.695	0.007
Earnings Margin	0.014	0.000	0.000	0.618	0.536
Debt/ Equity	0.080	0.352	0.099	3.551	0.000
Volatility	-0.057	-0.007	0.003	-2.537	0.011

This model has the highest fit of those tested so far, with an R^2 of 0.326. Earnings margin, a performance measure, is insignificant, although volatility (unusually) is significant. Like the first version of regression 2, this one shows size to be the most important variable by a large margin, in this case by an even larger margin. Only it and US listing show non-negligible standardised coefficients. Looking to residuals shows clear problems, however; the pattern appears largely unchanged from the original regression 2.

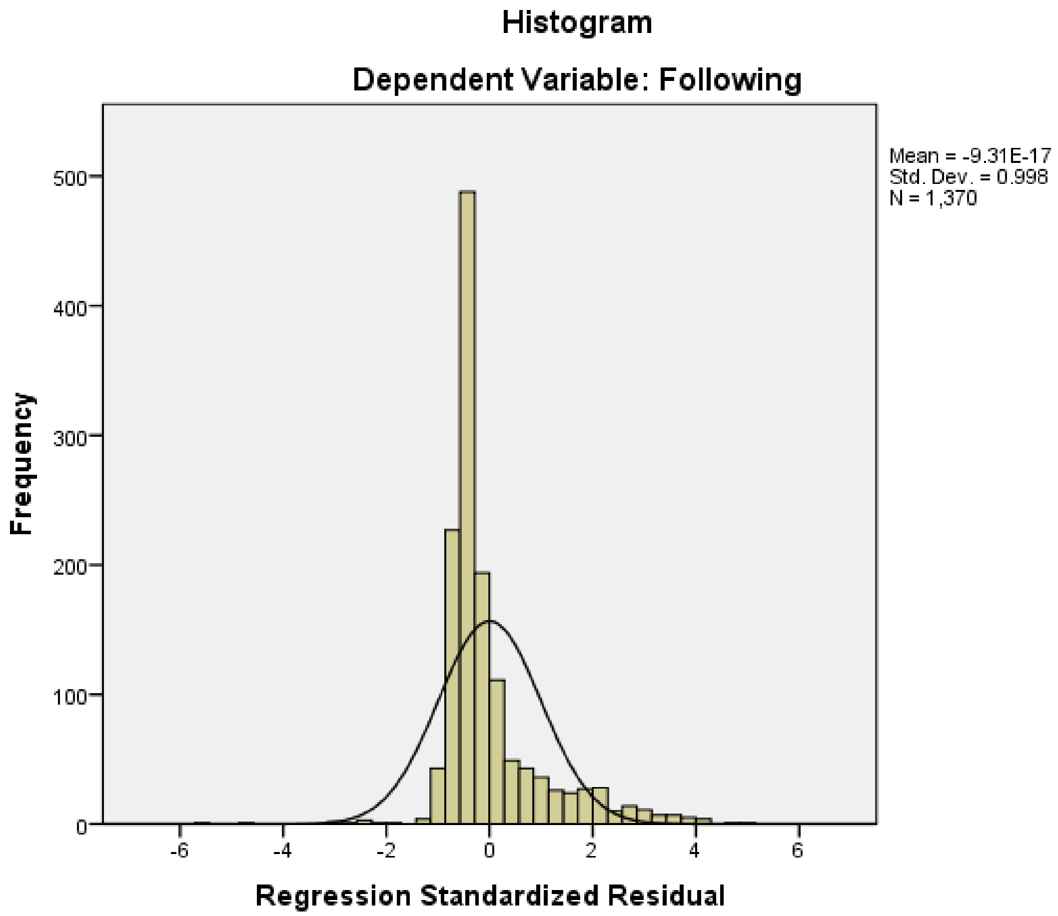

Figure 6.5: Residual Histogram of Regression 2 with Outliers Removed

6.3.3: Regression 3

Regression 3 retested without outliers produces results unlike the original version.

Table 6.6: Results of Regression 3 with Outliers Removed					
Model Data	F-stat	Model Sig.		R-square	Adj. R-square
	121.599	0.000		0.349	0.346
Variable Data	Std. Coeff.	Coeff.	Std. Error	t-stat	Sig.
Constant		1.863	0.233	7.996	0.000
Market Cap	0.296	0.000	0.000	9.429	0.000
Foreign listing	0.269	2.223	0.189	11.767	0.000
Sensitive	0.044	0.554	0.286	1.938	0.053
Profit/Size	0.215	4.205	0.447	9.398	0.000
Debt	0.124	0.000	0.000	3.956	0.000
Volatility	-0.003	0.000	0.003	-0.129	0.898

The R^2 here of 0.346 is even higher than the no-outlier regression 2 above. Where the original version of this model showed listing to be very important, it now has the second largest standardised coefficient behind market cap. Debt was formerly third, but now comes behind profit/size. Unusually, the sensitivity variable is narrowly insignificant here.

Residual plotting shows improvement over the initial version of this model, but does not indicate that non-normality of residuals has been eliminated by removing outliers.

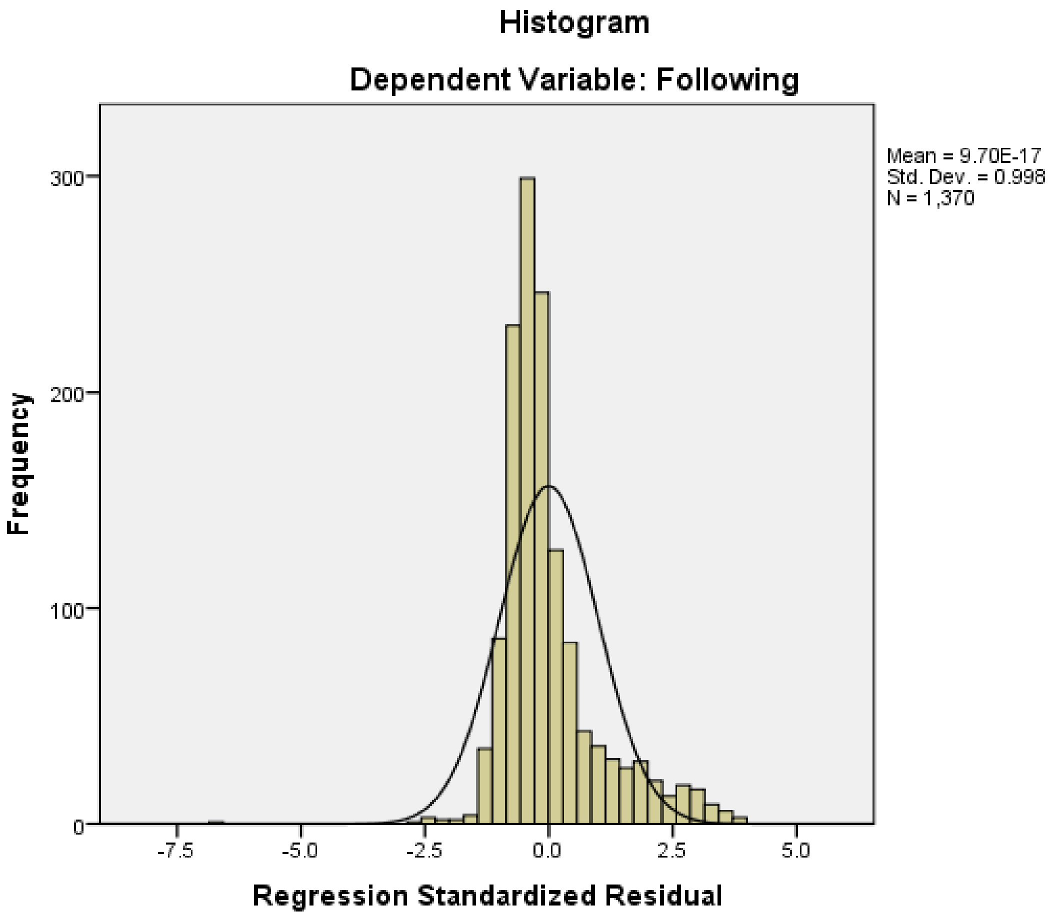

Figure 6.6: Residual Histogram for Regression 3 with Outliers Removed

Across the three models in this section, little information on disclosure theory emerges. Model 3 has the best fit and an insignificant sensitivity variable, showing little support for Legitimacy explanations. Model 2 has a slightly lower fit and an insignificant performance variable, not supporting a Signalling explanation. This would suggest Agency Theory is the best explanation by default.

However, while their R^2 values are much larger than that of model 1 above, model 1 has the most normal residuals. It explains less of the variance in the disclosure measure, but better follows regression assumptions. These three models are therefore considered weak evidence for Agency Theory.

6.4: Logarithmic data regressions

Logarithmic data transformations improve the R^2 and residual normality of all models tested.

6.4.1: Regression 1

Table 6.7: Results of Regression 1 with Logarithmic Data					
Model Data	F-stat	Model Sig.		R-square	Adj. R-square
	319.144	0.000		0.573	0.571
Variable Data	Std. Coeff.	Coeff.	Std. Error	t-stat	Sig.
Constant		93.744	23.555	3.980	0.000
Book Value	0.728	1.830	0.052	35.490	0.000
Foreign listing	0.103	0.885	0.166	5.316	0.000
Sensitive	0.038	0.512	0.239	2.141	0.032
Return on Assets	-0.086	-30.652	6.575	-4.662	0.000
Debt/ Assets	0.035	0.131	0.067	1.947	0.052
Volatility	-0.015	-0.051	0.058	-0.875	0.382

Regression 1 shows R^2 of 0.571 when logarithmic data is used. Book value, the size measure, is the most important variably by a large margin. Volatility is firmly insignificant, while debt/assets is marginally insignificant. This model suggests a rejection of Agency explanations as a result, but the low standardised coefficients on both the sensitive and performance measures suggest little support for either. Residuals from the model are marginally skewed and noticeably leptokurtic, although not to the extent of models examined so far.

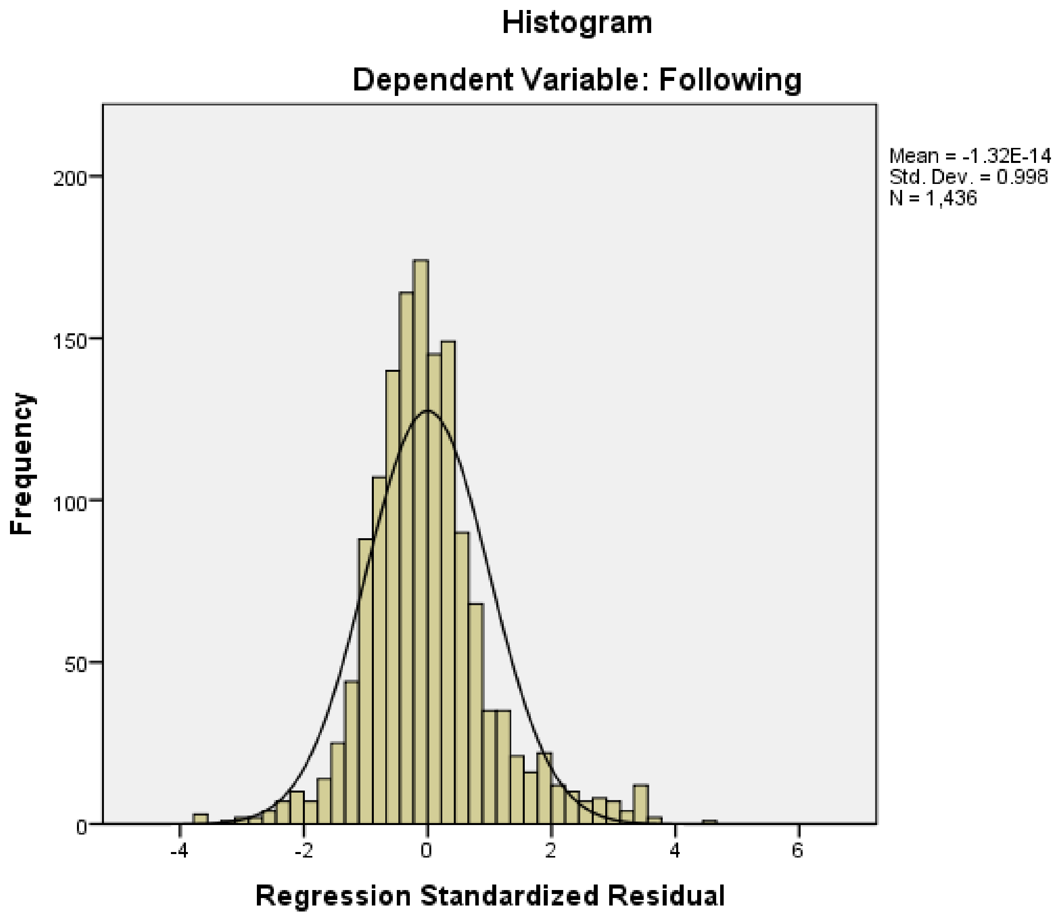

Figure 6.7: Residual Histogram for Regression 1 with Logarithmic Data

6.4.2: Regression 2

Table 6.8: Results of Regression 2 with logarithmic data					
Model Data	F-stat	Model Sig.		R-square	Adj. R-square
	400.543	0.000		0.627	0.626
Variable Data	Std. Coeff.	Coeff.	Std. Error	t-stat	Sig.
Constant		-22.239	4.106	-5.417	0.000
Revenue	0.767	2.444	0.053	46.465	0.000
US listing	0.062	1.055	0.279	3.784	0.000
Sensitive	0.084	1.121	0.219	5.121	0.000
Earnings Margin	-0.007	-0.131	0.322	-0.406	0.685
Debt/ Equity	0.002	0.009	0.062	0.152	0.879
Volatility	0.004	0.014	0.055	0.264	0.792

Regression 2 with logarithmic data has the single highest R^2 in the thesis, with a value more than double those found when using untransformed data (0.626). Revenue, the size measure, is the most powerful variable by a large margin, and of the three the indicate theories only sensitivity is significant. This suggests support only for a Legitimacy explanation of disclosure.

However, while the R^2 is high, the residual plot below shows clear skew and kurtosis. As in the outlier-free data, the best fit is not necessarily indicative of best meeting regression assumptions.

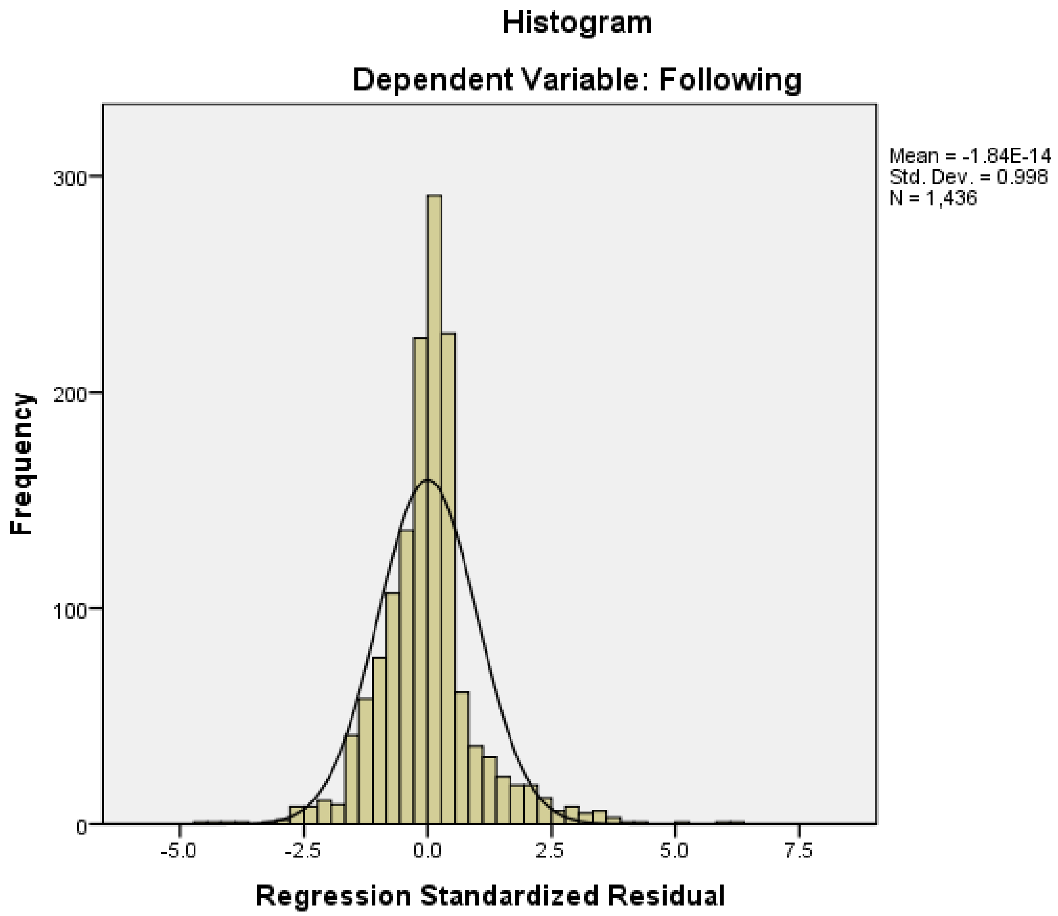

Figure 6.8: Residual Histogram for Regression 2 with Logarithmic Data

6.4.3: Regression 3

Table 6.9: Results of Regression 3 with logarithmic data					
Model Data	F-stat	Model Sig.		R-square	Adj. R-square
	266.874	0.000		0.528	0.526
Variable Data	Std. Coeff.	Coeff.	Std. Error	t-stat	Sig.
Constant		10.935	42.868	0.255	0.799
Market Cap	0.461	1.001	0.046	21.578	0.000
Foreign listing	0.173	1.485	0.171	8.704	0.000
Sensitive	0.071	0.955	0.250	3.817	0.000
Profit/ Size	-0.009	-4.501	9.132	-0.493	0.622
Debt	0.279	0.363	0.026	13.754	0.000
Volatility	-0.039	-0.133	0.061	-2.166	0.030

Logarithmic regression 3 shows the lowest R^2 of the logarithmic models, although still higher than the untransformed or outlier-free models at 0.526. The size variable is still the most important, although by a smaller margin than in the other log models. Only profit/size is insignificant, rejecting a Signalling explanation while supporting both others. The high standardised coefficient of debt suggests this model provides more support for Agency Theory than Legitimacy.

Residual analysis shows more normality than those of regression 2, but less than regression 1.

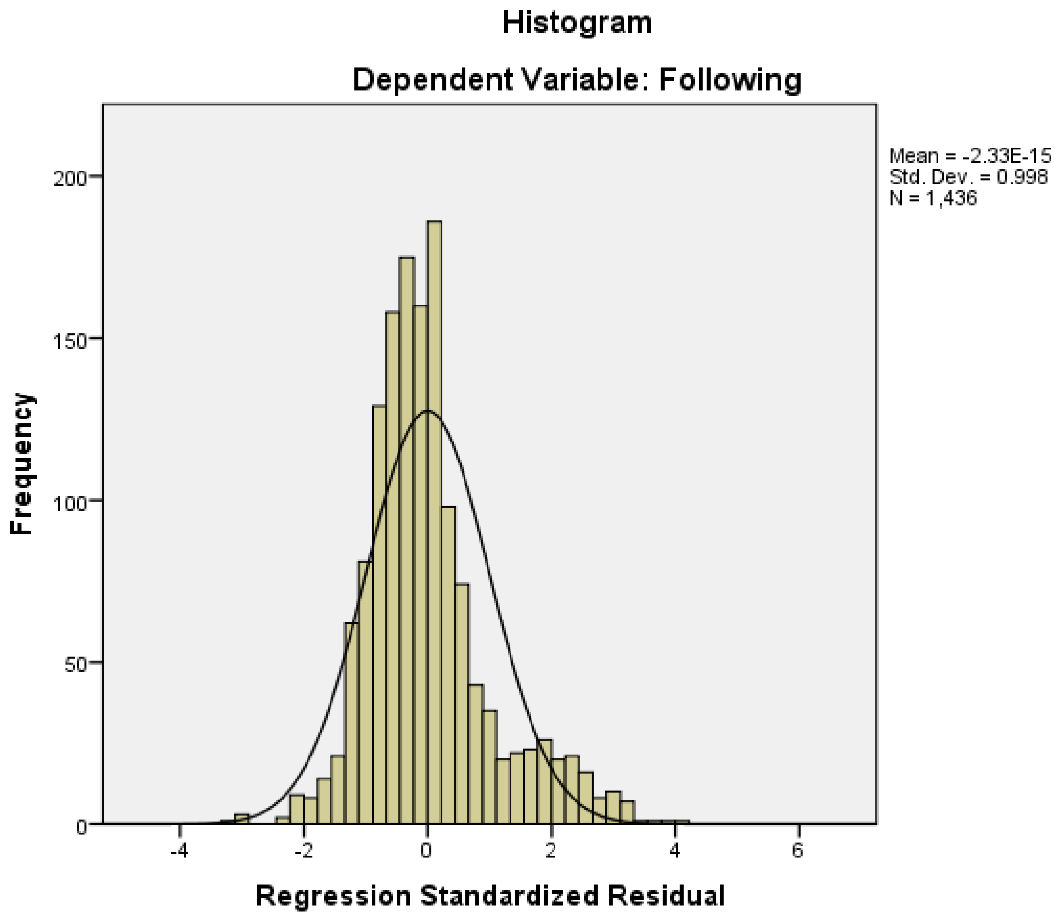

Figure 6.9: Residual Histogram for Regression 3 with Logarithmic Data

The logarithmic data does not present a theory as the best supported. Regression 1 rejects Agency due to the insignificant debt finance variable, but shows little support for either alternative due to the low standardised coefficients. Regression 2 only supports Legitimacy Theory, while regression 3 rejects Signalling while providing some support for the others. This overall suggests limited support for Legitimacy as the best explanation of disclosure as it is consistently not rejected.

6.5: Normal Score Regressions

The use of normal score transformations greatly improves the normality of residuals. Logarithmic data is better for R^2 values, but normal scores still provide a large improvement.

6.5.1: Regression 1

Table 6.1: Results of Regression 1 with Normal Score Data					
Model Data	F-stat	Model Sig.		R-square	Adj. R-square
	262.098	0.000		0.524	0.522
Variable Data	Std. Coeff.	Coeff.	Std. Error	t-stat	Sig.
Constant		2.883	0.209	13.762	0.000
Book Value	0.591	3.942	0.163	24.160	0.000
Foreign listing	0.152	1.304	0.177	7.352	0.000
Sensitive	0.066	0.884	0.254	3.479	0.001
Return on Assets	0.085	0.564	0.144	3.925	0.000
Debt/ Assets	0.019	0.140	0.139	1.002	0.316
Volatility	0.000	0.001	0.128	0.011	0.991

This model's R^2 of 0.522 is among the higher in this thesis, but lower than those of models using logarithmic data. Company size (book value) is by a large margin the most important variable. Like in log regression 1, the debt finance measure is insignificant, rejecting Agency Theory, and the standardised coefficients of the sensitivity and performance measures are low enough to provide little support for their respective theories.

While still slightly skewed and leptokurtic, the residuals of this model are among the best in the thesis.

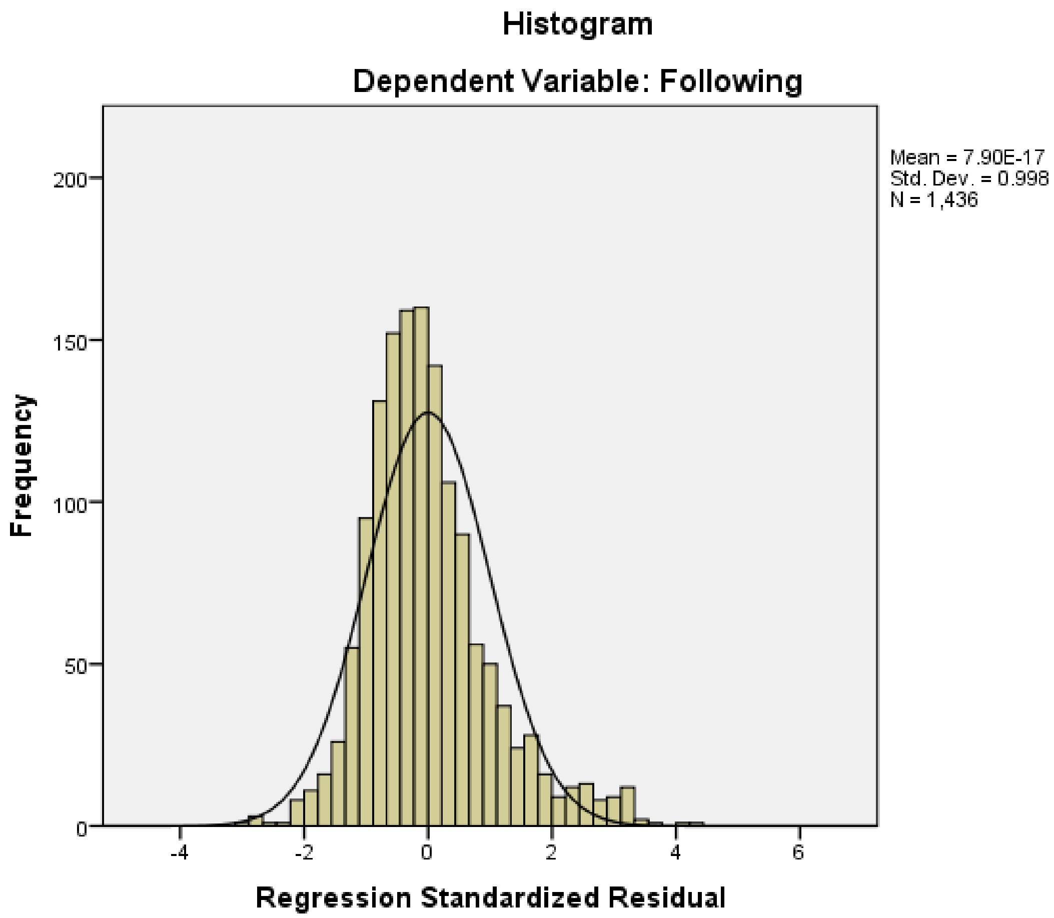

Figure 6.10: Residual Histogram for Regression 1 with Normal Score Data

6.5.2: Regression 2

Table 6.11: Results of Regression 2 with Normal Score Data					
Model Data	F-stat	Model Sig.		R-square	Adj. R-square
	258.137	0.000		0.520	0.518
Variable Data	Std. Coeff.	Coeff.	Std. Error	t-stat	Sig.
Constant		3.023	0.175	17.253	0.000
Revenue	0.709	4.832	0.160	30.255	0.000
US listing	0.089	1.503	0.316	4.762	0.000
Sensitive	0.162	2.167	0.250	8.675	0.000
Earnings Margin	-0.009	-0.061	0.142	-0.427	0.669
Debt/ Equity	-0.045	-0.311	0.144	-2.157	0.031
Volatility	0.038	0.038	0.130	0.293	0.769

At 0.518, this model has the lowest fit of the three normal scores models, which is unusual for regression 2. Standardised coefficients demonstrate revenue (size) to be the most important variable by a large margin, followed by sensitivity, best supporting Legitimacy Theory. The low (absolute) value of the debt finance variable's standardised coefficient suggests neither support nor rejection of Agency Theory. Earnings margin is insignificant, rejecting a Signalling explanation.

The residuals of this model are not as Normal as those of Normal Score Regression 1, but better than most in the thesis. Skewness is more apparent in this diagram than that of Regression 1 above.

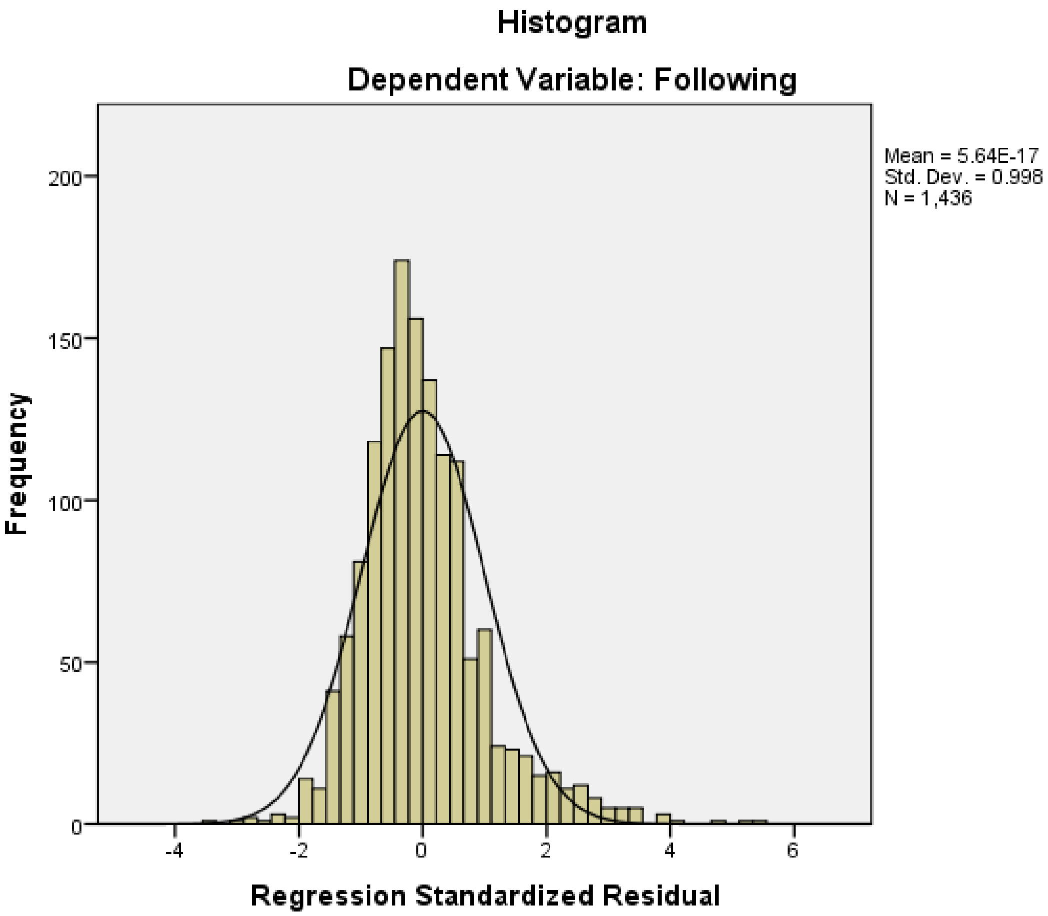

Figure 6.11: Residual Histogram for Regression 2 with Normal Score Data

6.5.3: Regression 3

Table 6.12: Results of Regression 3 with Normal Score data					
Model Data	F-stat	Model Sig.		R-square	Adj. R-square
	344.346	0.000		0.591	0.589
Variable Data	Std. Coeff.	Coeff.	Std. Error	t-stat	Sig.
Constant		2.862	0.191	15.019	0.000
Market Cap	0.454	3.056	0.154	19.792	0.000
Foreign listing	0.141	1.204	0.162	7.415	0.000
Sensitive	0.070	0.941	0.234	4.012	0.000
Profit/Size	0.080	0.531	0.127	4.192	0.000
Debt	0.285	2.064	0.149	13.878	0.000
Volatility	-0.016	-0.110	0.118	-0.933	0.351

At 0.589, Regression 3 is the best of the Normal Scores model. As with the logarithmic data, the market cap is by far the most important variable, albeit to a lesser extent than size measures have been in other models. Standardised coefficients show importance of the debt finance variable, supporting Agency Theory above other theories.

Residuals of this model are comparable to those of Normal Score Regression 2, being good but less so than those of Regression 1 with the same data.

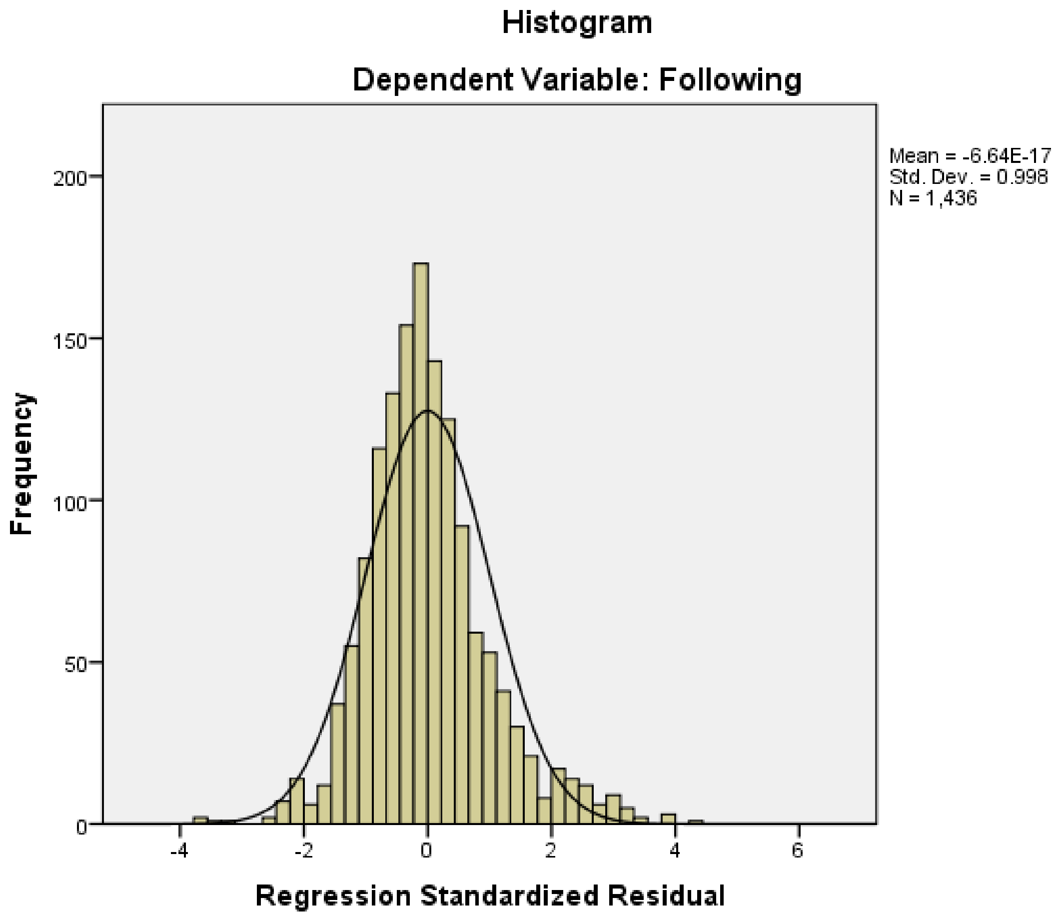

Figure 6.12: Residual Histogram for Regression 3 with Normal Score Data

A clear best supported theory emerges from regressions performed using Normal Score data. Legitimacy Theory is supported by models 1 and 2, while model 3 does not reject it.

6.6: Overall Regression Discussion

Table 6.13 shows the adjusted R^2 of all models and the standardised estimates of each variable. Estimates of NS indicate that the value was not significant at the 5% level.

	Table 6.13: Fit and estimates for all regression models							
	Fit		**Standardised Estimates**					
Model	**R-square**	**Adj. R-square**	**Size**	**Listing**	**Sensitive**	**Performance**	**Debt**	**Volatility**
Base Data Models								
Reg 1	0.233	0.230	0.204	0.343	0.077	0.161	0.064	NS
Reg 2	0.126	0.122	0.256	0.148	0.140	NS	NS	NS
Reg 3	0.214	0.211	0.092	0.350	0.075	0.066	0.168	NS
Outlier-free Models								
Reg 1	0.281	0.277	0.246	0.289	0.068	0.255	0.099	NS
Reg 2	0.329	0.326	0.522	0.114	0.061	NS	0.08	-0.057
Reg 3	0.349	0.346	0.296	0.269	NS	0.215	0.124	NS
Logarithmic Models								
Reg 1	0.573	0.571	0.728	0.103	0.038	-0.086	NS	NS
Reg 2	0.627	0.626	0.767	0.062	0.084	NS	NS	NS
Reg 3	0.528	0.526	0.461	0.173	0.071	NS	0.279	-0.039
Normal Score Models								
Reg 1	0.524	0.522	0.591	0.152	0.066	0.085	NS	NS
Reg 2	0.520	0.518	0.709	0.089	0.162	NS	-0.045	NS
Reg 3	0.591	0.589	0.454	0.141	0.070	0.080	0.285	NS

Taken collectively, the regression models of all types and using all data most support Legitimacy Theory as an explanation of disclosure. Models using the unaltered, logarithmic, and normal score data all support this theory above any other. The models using the untransformed data with outliers removed generally support Agency Theory. However, these are merely the most consistently supported theories among the regressions using that form of data. The other theories may be rejected by some models, but are generally supported to a smaller extent. Signalling Theory and Agency Theory are less consistently supported or less powerful across all models, but each demonstrate some cases where they best explain disclosure as seen among the sample.

Size, by all measures used in the regressions above, has a uniformly positive effect on disclosure. Such a result is to be expected based on the literature to date. Size was one of the first variables identified as a potential explanation. It had a positive effect when Singhvi and Desai (1971) used it, and the paper heavily criticising their work argues that size should be even more important (Buzby, 1975). The vast majority of papers examined include size and find it to be a positive determinant of disclosure. The identified exceptions invariably find an insignificant result and never negative outcomes. Most of the exceptions are studying something more specific than the generalised studies that find positive results. For example, Malone et al (1993) offer an early insignificant size finding in a sample consisting only of firms in the oil and gas sector. Frankel et al (1994) uses a more general sample, but the dependent variable is the existence of a voluntary manager-provided forecast of results instead of a more general measurement of disclosure. More recently, Ali et al (2007) and Chen et al (2008) have had insignificant size results when examining family-owned companies in the USA.

Listing status, measured above as multiple listing, is invariably positive in the regressions above. This too is in keeping with literature to date. Multiple listing does not have the same frequency of occurrence in literature as size, but is positive in a majority of cases. Taylor et al (2010) find a rare negative result, but believe this apparent oddity is considered a unique result of their combined sample and disclosure measure. Raffournier (1995) and Robb et al (2001) both find insignificant multiple listing, neither of which are easily explained.

Sensitive industry membership has a near-uniform effect, being insignificant in one model (logarithmic 2) but otherwise positive. The effects of various industry classifications have been tested in the literature and tend to have any effects that the researcher predicted in advance. Adams et al (1998) have industry effects comparable to those seen here; their sample is split into categories that are expected to face different levels of public scrutiny, finding that those expected to be more sensitive to public opinion tend to disclose more than others.

The various measures of financial performance have a more mixed set of results. The obtained coefficient is significant and positive in seven models, but the standardised coefficients suggest a weak overall effect in many cases. Regression 3 has particularly mixed results here, having a weak performance effect with the base and normal data, a stronger effect with outliers removed, and an insignificant effect with logarithmic data. In wider literature, performance measures have had an inconsistent effect on disclosure, being positive about as often as negative and insignificant in about as many cases as significant. There is little pattern to the wider results in literature, with all three possibilities occurring in various contexts and time periods.

Debt financing measures have very mixed results. Regression 1 uses Debt/Assets, finding a positive but weak result with the base and outlier-free data, but insignificant results with the two transformed data types. Regression 2, using Debt/Equity, has an insignificant coefficient with the basic and logarithmic data, a weak positive with the outlier-free data, and a weak negative with normal score data. Regression 3 has a clear positive result in most cases, but the debt measurement in this case is simply the value of debt in the firm, not scaled by size. This was demonstrated in Chapter 5 to be strongly correlated with size, so it is likely that this measure is actually indicative of size, despite finding a low VIF in testing.   Like performance, the effects observed from debt finance have varied greatly in the literature and has a similar set of results – positive as often as negative, significant as often as not. There is, however, a small pattern within the findings on this variable. The positive results identified have clustered towards earlier papers (Malone et al, 1993, Frankel et al, 1995, and Ahmed and Courtis, 1999) and cease appearing by 2000. However, with only three positive results identified, this may not be a meaningful pattern and further evidence would be needed to reach a firmer conclusion.

Volatility is here found to be insignificant overall, being firmly insignificant in all but two models. Findings in literature are quite evenly split between positive, negative, and insignificant effects on disclosure, and like performance there is little apparent pattern. There are some signs of negative results being a more recent event as the first identified paper with such is from 2003 (Field et al), but as with debt this is a conclusion drawn from a few observations and more are necessary to confirm the effect.

On the whole, the results obtained here are in line with past research. Size and listing are both consistently positive, as found here. Industry effects are commonly significant, albeit not in any one form, consistent with the result here. Recent performance and debt finance are unclear, having mixed results with a tendency towards weak positives here, and do not have any consistent effects in literature. Finally, volatility is not significant, which is overall in keeping with the literature's even mix of results.

6.7: Structural Models of Regression

This section covers the results of fitting structural models that match the three regressions above. This is done to test for comparability of the methods.

Such simple models do not take full advantage of the power of SEM. By their nature as regressions, none of these models contain multi-indicator latent variables, instead having only single indicators. This removes the measurement model aspect in favour of the assumption that everything is measured appropriately and gives no means of checking that this is correct. In addition, the validity and reliability tests for SEM rely upon the existence of latent variables, meaning this step has to be skipped for these models. Further, these models follow the regression assumption of independence among explanatory variables, eliminating another potential advantage of SEM.

The results demonstrate that the two methods are comparable in terms of the findings. In all cases, regardless of data type, the resulting SEM are nearly identical to the original regressions. Only the models using the base, unaltered data are reported here in order to illustrate the comparability without repeatedly reporting a model that has effectively been tested earlier in the chapter.

Visually, these models are all similar to:

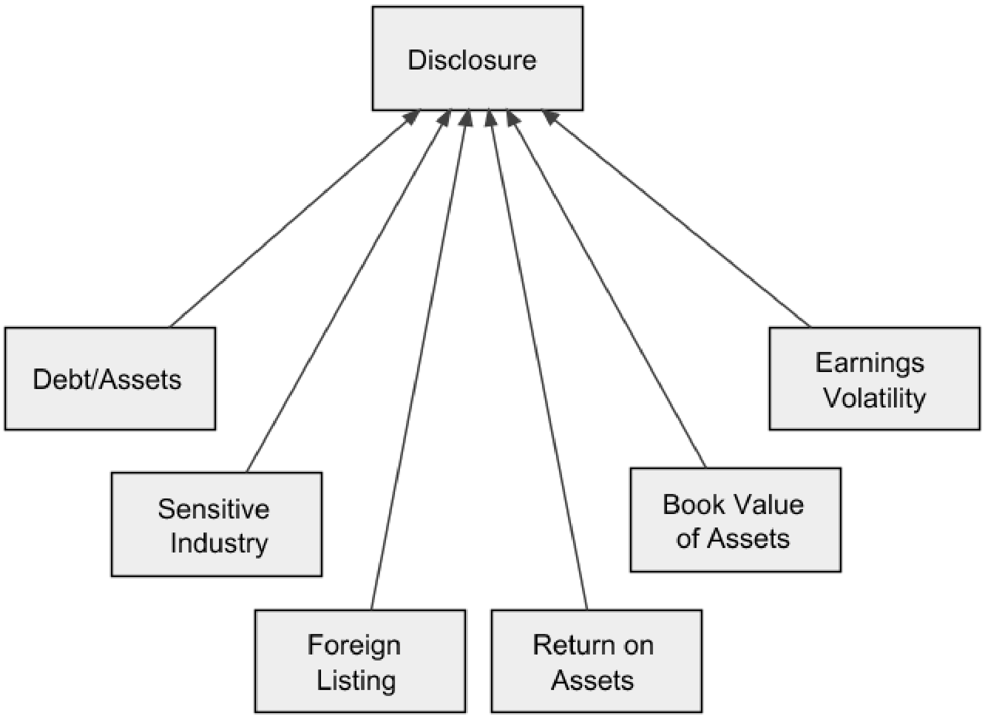

Figure 6.13: Example of Regression as SEM

This diagram represents Regression 1. The only changes any other Regression-type models make are alternative variables, as discussed throughout section 6.2. Each table of results shows six items, each of which is a single variable with an arrow pointing towards analyst following.

For each model, a table of fit values is produced. SEM lacks a single convenient measure of fit like R^2 and a range of measures are used instead for a rounded view of how each model fits. Each model's fit table, tables 6.14, 6.16, and 6.18, contain the measurements discussed below.

Chi-square is the value of the chi-square statistic obtained for the model. It is not especially useful in itself and is included for completeness of information rather than its value as an indicator of fit. Good fit is indicated by a low value relative to the next measurement.

DF is the degrees of freedom in the model, and again is included for information rather than any particular testing value. Degrees of freedom in SEM are based on the number of parameters in the model and not the sample size. The calculation involves the number of covariances between parameters, which for p parameters is $(p(p+1))/2$. The number of covariances that are given a fixed value in the model is then subtracted from this value. In the regression models in this section the calculation is consistent. Each model has seven total variables for a total of (7*8)/2 = 21 total possible covariances. However, the assumption of variable independence means many of these are assigned a value of 0 rather than being estimated. Ultimately, only six are estimated, being the covariances between the independent variable and each of the six explanatory variables, so the DF value for each model is 21 – 6 = 15. AMOS (one of the software packages used) explains the calculation a little differently. It instead suggests there are 28 distinct sample moments, 21 covariances as above plus 7 means. Of these, only the 6 covariances not set to 0 and the 7 means are estimated, leading to 13 estimates and a DF calculation of 28 – 13 = 15.

P-value is the probability of the obtained chi-square value occurring on the mentioned number of degrees of freedom. Good fit is, unusually, indicated by a high p-value. The test used has good model fit as its null hypothesis, so a significant result indicates a large deviation from an assumption of good fit. Mathematically, the process checks whether a difference between model and observation is significant. A high p-value therefore indicates an insignificant difference, meaning that in an unusual variation on normal testing, a significant result indicates poor fit. All other references to significance in this document use the 5% level, so for consistency this will be applied here as well; a p-value of 0.05 or higher indicates good fit. Further, this test has problems with large samples; if enough data points are tested, the cumulative total discrepancy becomes large enough to be significant and suggest a poor fit even if the individual differences are very small.

CMIN/DF is the obtained chi-square value divided by the degrees of freedom. This gives a quick but imprecise view of fit and serves as a useful rough measure. For ideal fit, this ratio should be as low as possible. In practice, values of 1 are excellent, 2 is good, and 3 indicates a mostly-fitting model with some problems. This measure is affected by the sample size issue of the chi-square test above as it serves as the numerator in the function.

NFI, and TLI are fit indices. In each case, the measures take values in the range [0, 1], where 1 indicates perfect fit and 0 means the model explains nothing (both can mathematically go beyond the range, but this requires extremely good or bad fit to occur). These measures fit relative to another, theoretical model which is not designed to fit well. In all cases, values near 1 are ideal; as a general rule, values of 0.9 and above are definite indicators of good fit.

PNFI is a parsimony-adjusted version of the NFI above. This means that it penalises complexity in the model. It is possible with SEM to make a model that has plenty of linkages and complicated relationships between variables and find that this fits well because it is mapping the sample near-perfectly, leading to a model that has severely limited generalizability. Parsimony adjustment alters the formula for the index to include some aspect that reduces the value for everything added to the model, so inclusions will only be beneficial if their addition to fit outweighs the penalty term they apply. Note that the parsimony adjustment is applied to all models, so even a simple one has a lower fit under this index than the unadjusted NFI. Even a simple model will therefore have a lower fit under this measure than a non-adjusted index. As a result, the judgements made on this index must be more lenient; a value above 0.9 is very hard to obtain. 0.7 is used instead as the cut-off of clear powerful fit.

RMSEA is the Root Mean Square Error of Approximation, a complex but reliable measurement of model fit. RMSEA below 0.08 is a definite indication of good fit, while 0.05 is excellent.

The final column, B-S P-value, is the result of testing a Bollen-Stine bootstrap method on the model. This is the probability AMOS (the SEM software used) has determined of the model being correct from using a bootstrapping approach. This is essentially a significance test that works the same way as the other P-value column in that the null hypothesis is that the model is correct and an insignificant result is therefore indicative of good fit. The reason for its inclusion is that the bootstrap method involved corrects for non-normality of the data, which is an important factor in the sample employed here.

6.2.1: Regression 1

This model takes the basic regression equation described just under the Regression 1 heading above and turns it into a structural model. Everything that is listed on the right-hand side is considered as a cause of the disclosure measure of analyst following, all independent of each other.

Table 6.13: Regression 1 SEM estimates

			Estimate	Standard error	C.R.	P-value	Standardised Estimate
Following	<---	Book Value	0.000	0.000	8.839	0.000	0.208
Following	<---	Foreign	2.934	0.198	14.817	0.000	0.349
Following	<---	Sensitive	1.023	0.309	3.309	0.000	0.078
Following	<---	Return on Assets	1.700	0.244	6.967	0.000	0.164
Following	<---	Debt/ Assets	0.416	0.151	2.756	0.006	0.065
Following	<---	Volatility	0.000	0.001	-0.518	0.605	-0.012

When compared to table 6.1, it becomes clear that the two methods do not differ greatly in results. All estimates, both standardised and not, are nearly identical across the two.

While these tables show some usefulness to the model, the overall fit is not acceptable in any regard and is poor by several measures.

Table 6.14: Regression 1 SEM fit

Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
318.392	15	0.000	21.226	0.544	0.374	0.389	0.119	0.139

The RMSEA is too high for a definite good fit. However, even the basic indices indicate poor fit here, which is unusual when compared against models later in the thesis. The parsimony-adjusted indices are actually higher than a few later models, which is likely a result of the extreme simplicity of these remade regressions.

However, this is all consistent with the fit obtained earlier. The model as a regression was found to have R^2 of 0.230, a low figure that falls below even the usually low figures obtained in similar research. A poor fit for this SEM is to be expected.

6.2.2: Regression 2

When analysed as SEM, the estimates for this models are very close to those of the regression version in section 6.2.2, including a number of insignificant items:

Table 6.15: Regression 2 SEM estimates

			Estimate	Standard error	C.R.	P-value	Standardised Estimate
Following	<---	Revenue	0.000	0.000	10.376	0.000	0.258
Following	<---	US	2.513	0.418	6.008	0.000	0.150
Following	<---	Sensitive	1.867	0.330	5.656	0.000	0.141
Following	<---	Earnings Margin	0.000	0.000	0.410	0.682	0.010
Following	<---	Debt/ Equity	0.001	0.030	0.043	0.850	0.966
Following	<---	Volatility	0.000	0.000	-.312	0.755	-0.042

This is the best-fitting of all the regression remakes, which goes against the very poor fit of 0.125 this model had as a regression. Index-type fit measures are still too low to be clearly acceptable, but this model passes the RMSEA test with a value of 0.049.

Deletion of the insignificant items was tested. This improves the index measures, but increases the RMSEA to 0.085, just above the acceptable level. The confidence interval of this estimate goes below the required level, giving some indication that it may be acceptable.

Table 6.16: Regression 2 SEM fit								
Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
67.262	15	0.000	4.484	0.741	0.694	0.529	0.049	0.070

Despite the simplicity of this model, the PNFI is still low. The B-S p-value is also very low, albeit not quite enough to reject the null hypothesis of this model being correct. This is consistent with the comparable regression having a notably lower fit than the others.

6.2.3: Regression 3

Like the previous two models, this SEM's estimates largely match those of the relevant regression model:

Table 6.17: Regression 3 SEM estimates							
			Estimate	Standard error	C.R.	P-value	Standardised Estimate
Following	<---	Market Cap	0.000	0.000	3.927	0.000	0.094
Following	<---	Foreign	2.997	0.200	14.952	0.000	0.358
Following	<---	Sensitive	1.000	0.313	3.194	0.001	0.076
Following	<---	Profit/Size	0.299	0.106	2.817	0.005	0.067
Following	<---	Debt	0.000	0.000	7.184	0.000	0.172
Following	<---	Volatility	-0.001	0.001	-1.209	0.227	-.029

The fit here is essentially not quite as good as regression 2 in any regard.

Table 6.18: Regression 3 SEM fit								
Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
166.880	15	0.000	11.125	0.673	0.567	0.482	0.084	0.174

Despite being one of the best fitting when tested as a regression, the SEM version does not have an absolutely good fit. There is little to recommend this model. The closest it comes to good fit is that the RMSEA narrowly fails to reach the acceptable level, while even the basic indices are clearly too low.

6.2.4: Additional Models

Retesting regression models in SEM form produces near-identical coefficient estimates in all cases tested here and detailed discussion of most models is not provided. Model fits are quite different, however. Table 6.19 below presents the fit for all regression models retested as SEM.

Models are named according to the data type used, with base referring to untransformed data, outlier to that with outliers removed, log to logarithmic data, and norm to normal score data. The number in each name defines whether the model is regression 1, 2, or 3 as above.

Table 6.19: Fits for All Regression Structural Models

	Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	BS P-value
Base 1	318.392	15	0	21.226	0.544	0.374	0.389	0.119	0.139
Base 2	67.262	15	0	4.484	0.741	0.694	0.529	0.049	0.070
Base 3	166.880	15	0	11.125	0.673	0.567	0.482	0.084	0.174
Outlier 1	292.420	15	0	19.495	0.607	0.462	0.433	0.116	0.005
Outlier 2	148.120	15	0	9.875	0.786	0.723	0.562	0.081	0.005
Outlier 3	1245.920	15	0	83.061	0.320	0.049	0.229	0.245	0.005
Log 1	622.404	15	0	41.494	0.662	0.533	0.473	0.168	0.005
Log 2	177.939	15	0	11.863	0.889	0.855	0.635	0.087	0.005
Log 3	618.686	15	0	41.246	0.635	0.496	0.454	0.167	0.005
Norm 1	1027.79	15	0	65.519	0.509	0.316	0.363	0.217	0.005
Norm 2	909.836	15	0	60.656	0.537	0.355	0.383	0.204	0.005
Norm 3	1244.42	15	0	82.961	0.508	0.313	0.363	0.239	0.005

In table 6.13 (in section 6.6), there is a clear trend in the R^2 and adjusted R^2 fit measures. The base data tends to produce the lowest values and removing outliers improves things to some extent. Normal score data provides a larger improvement and logarithmic data the greatest improvement, with many of the fit values for these models being very high.

Table 6.19 above does not show the same clear changes in fit values when the data type is changed, however. While the logarithmic data does tend to produce the best fits by the index measures (NFI, TLI, and PNFI), it is not as clear as the improvement to R^2 for the same models. The normal score data often shows worse fit than the base data in these models.

6.8: Conclusion

There are two questions central to this chapter. Are the methods comparable? What do the models imply?

In answer to the first question, the methods obtain similar estimates. Table 6.1 and 6.13 show the estimated coefficients for each variable in regression 1 and the same model remade as a SEM, respectively. Comparing the standardised estimates column in each table reveals only minor differences. In most cases, the values differ only at the third decimal place, suggesting only minor differences. As this is the final digit given in each table, this may be the result of a rounding process in the calculations rather than an actual difference. Foreign listing is the exception, having a standardised estimate of 0.343 under regression and 0.349 in SEM, a larger discrepancy but still a minor difference. Comparison of p-values reveals similarly small differences between the methods. The same pattern is apparent with the other two models and with the unreported structural models using the other data types. Given the same model, the two approaches produce almost identical estimates and drawn the same conclusions about variable significance, supporting the idea that the two are comparable.

One difference between methods is apparent, however. Table 6.13 shows a large improvement in R^2 when the logarithmic or normal score data is used in regressions. This is not the case for SEM, as shown in table 6.19 and discussed in section 6.2.4 above. In this regard, regression has an advantage over the more complex SEM.

The implications of these models are largely discussed in section 6.6 above. Size and multiple listing status each have positive effects on disclosure in all three models, as expected from the literature. More useful is a comparison of the financial performance, debt finance, and sensitive industry membership variables, which are respectively the primary indicators of Signalling Theory, Agency Theory, and Legitimacy Theory. Performance and debt are each significant in regressions 1 and 3, indicating some support for Signalling and Agency explanations. However, sensitivity is significant in all three cases, suggesting that Legitimacy is a more robust explanation as it applies in all three models. All three explanations are supported to some extent, with Legitimacy having marginally more support than the others.

However, this is a conclusion drawn only by examining the implications of generalised models. The next chapter covers the use of SEM to make models specifically to represent each theory, which may offer a different view on the theories studied.

Chapter 7: Theory Modelling

This chapter covers the modelling of theories using a SEM approach. The four sections to the chapter are described below.

Section 7.1 is a raw table of results. It is placed here for immediate reference.

Section 7.2 explains the preparatory steps taken before modelling can begin. To summarise, when using latent variables there is a requirement to test each to ensure they will work as intended. This process is explained in the section, followed by a somewhat naïve model that includes all of the gathered variables as an exploration of what may and may not be useful in later models.

Sections 7.3 through 7.5 are the most important, covering the results of building models for each theory and analysing these in SEM terms to determine which (if any) theories offer an explanation for disclosure among the companies that make up the sample. Section 7.3 shows the results of testing a model of Signalling Theory with all four data types, 7.4 does the same for Agency Theory, and 7.5 the same for Agency Theory.

Section 7.6 concludes the chapter, examining both what implications these models have for the theories examined as explanations of disclosure and other interpretations that may exist.

As in Chapter 6, significance is accepted at the 5% level. In addition, the modelling takes an iterative approach in which the least significant items are deleted in order to better focus on what is found to be relevant. Variables are only removed from models if they are insignificant at the 10% level, intended to leave room for further alterations.

7.1: Raw Results

This section is a table of results only. The various models are given more detailed treatment below. This section exists purely to allow easy comparisons between them. Note that the three regression-equivalents from section 6.2 (titled Reg 1, 2, and 3 here) are included for further comparison. The exploratory model rows represent the models in section 7.2 below, named such as these primarily exist to investigate the proposed latent variables for other models. The final three are each named for the theory they represent.

The various fit measures are as explained in section 6.7. A summary of each is presented here.

Chi-square is the value of the chi-square statistic obtained for the model. It is not directly useful as a measure of fit.

DF is the degrees of freedom in the model, and again is included for information rather than any particular testing value.

P-value is the probability of the obtained chi-square value occurring on the mentioned number of degrees of freedom. Good fit is indicated by a high p-value as the null hypothesis in this test is that the model fits the data well.

CMIN/DF is the obtained chi-square value divided by the degrees of freedom. This gives a quick but imprecise view of fit and serves as a useful rough measure. For ideal fit, this ratio should be as low as possible, with values below 3 being ideal.

NFI, TLI, and PNFI are fit indices. In each case, the measures take values in the range [0, 1], where 1 indicates perfect fit and 0 means the model explains nothing (both can mathematically go beyond the range, but this requires extremely good or bad fit to occur). These measures fit relative to another, theoretical model which is not designed to fit well. PNFI is the same as NFI but adds a penalty term for complexity in the model.

RMSEA is the Root Mean Square Error of Approximation, a complex but reliable measurement of model fit. RMSEA below 0.08 is a definite indication of good fit, while 0.05 is excellent.

The final column, B-S P-value, is the result of testing a Bollen-Stine bootstrap method on the model. This is the probability AMOS (the SEM software used) has determined of the model being correct from using a bootstrapping approach. Again, the null hypothesis is good fit.

Table 7.1: Table of SEM Results									
Model	**Chi-square**	**DF**	**P-value**	**CMIN/ DF**	**NFI**	**TLI**	**PNFI**	**RMSEA**	**B-S P-value**
Regression Models									
Reg 1	318.392	15	0.000	21.226	0.544	0.374	0.389	0.119	0.139
Reg 2	67.262	15	0.000	4.484	0.741	0.694	0.529	0.049	0.070
Reg 3	166.880	15	0.000	11.125	0.673	0.567	0.482	0.084	0.174
Exploratory Models									
Base	2177.565	24	0.000	90.732	0.841	0.764	0.561	0.250	0.005
Outlier	1180.702	65	0.000	18.165	0.899	0.865	0.642	0.112	0.005
Log	872.200	28	0.000	31.150	0.882	0.816	0.549	0.145	0.035
Normal	3122.814	46	0.000	67.887	0.791	0.703	0.551	0.216	0.050
Agency Theory Models									
Base	4.797	3	0.187	1.599	0.984	0.987	0.492	0.020	0.055
Outlier	918.821	54	0.000	17.015	0.885	0.891	0.637	0.108	0.010
Log	348.056	27	0.000	12.891	0.954	0.929	0.572	0.091	0.050
Normal	315.935	13	0.000	24.303	0.968	0.934	0.450	0.127	0.005

Table 7.1 continued : Table of SEM Results									
Model	**Chi-square**	**DF**	**P-value**	**CMIN/ DF**	**NFI**	**TLI**	**PNFI**	**RMSEA**	**B-S P-value**
Legitimacy Theory Models									
Base	179.117	16	0.000	11.195	0.986	0.977	0.563	0.084	0.557
Outlier	812.946	19	0.000	42.787	0.914	0.840	0.482	0.175	0.010
Log	288.977	22	0.000	13.135	0.961	0.940	0.587	0.092	0.085
Normal	436.835	20	0.000	21.842	0.957	0.926	0.532	0.121	0.005
Signalling Theory Models									
Base	180.239	23	0.000	7.836	0.986	0.980	0.631	0.069	0.562
Outlier	915.613	44	0.000	20.809	0.921	0.886	0.614	0.120	0.005
Log	312.755	28	0.000	11.170	0.958	0.938	0.596	0.084	0.005
Norm	1661.922	41	0.000	40.535	0.894	0.832	0.555	0.166	0.005

7.2: Exploratory Models

Each model is given the same series of analyses and resulting modifications. Before performing the SEM analysis, there are a number of preparatory steps. First, each latent variable must be tested for reliability and validity. Next, a confirmatory factor analysis (CFA) is run to test the measurement aspect of the model, before finally performing the SEM to test the structure side.

In each case, the pre-SEM analyses indicate a need to change plans. Some combination of the steps (usually validity and reliability) often indicated that latent variables were not working well. In many cases, the solution to this problem has been to replace latent variables with a single indicator. This suggests that a single measure is sufficient for many of the variables used in this field of research, possibly indicating that they are not complex and multifaceted concepts in need of complementary measurement methods.

Latent variables must be tested for reliability and validity, and a few tests are used. First, the average standardised loading for a given set of indicators must be 0.7 or above. Ideally, all indicators would load to this level on their own, but in some cases this is not possible. Following this, three reliability tests for latent variables are used. The first is Cronbach's (standardised) Alpha, which needs to be between 0.7 and 0.9 for a given set of indicators. Lower indicates that the variables are too divergent, while higher suggests that two or more are possibly identical measures. Second is Composite Reliability, a similar but more complex measure, which needs only to be above 0.7. The final measure is Average Variance Extracted (AVE), which works differently and needs only to be above 0.5.

Most important, the model is subjected to a confirmatory factor analysis (CFA) to test the measurement aspect. No CFA models are pictured here. In most cases, they are simply not useful to show as they are almost identical to the pictured SEM for the relevant case. The difference is that CFA is performed by removing all of the expected causal links between variables and instead allowing a covariance between every pair of variables.

The CFA process has been responsible for most of the changes made between initial models and the final tests. In each case, clear problems highlighted from the process are taken into account to form a new SEM, which is then subject to the same procedure. As expected, the new models rarely had any further problems highlighted by CFA. Another method of recommending changes to models also exists, known as the modification indices. After performing the calculations required for a SEM, these indices suggest possible relationships that could be added to the model in order to improve fit. They must be used with care, as the suggestions are based entirely on fit and not theory. However, after making CFA-derived changes, these indices largely suggested no further improvements.

Details on the process are given for a single model only, the one used to explore the intended latent variables. Others tended towards similar changes, making further discussion into repetition. This model is not based on a named theory. Instead, it simply assumes all data gathered is a potentially useful explanation of disclosure since each variable selected was based on existing research that contained justifications for it. The model uses all variables to attempt to explain disclosure. If a theory can be said to exist behind it, it is simply that this selection of variables may explain disclosure. This enables testing of the various planned latent variables without the influence of theory.

Volatility and sensitivity are each measured using single indicators in the regressions and are unchanged from this for SEM purposes.

Listing was originally a two-indicator latent, consisting of all foreign listing and US listing specifically. When both are used, problems consistently emerge involving the foreign listing measure. The latent variable has a standardised loading greater than 1 on foreign listing, which also carries a negative estimate of variance. The latter problem can sometimes be resolved in a simple manner; the estimate is negative, but the 95% confidence interval for it may range into acceptable positive numbers. This would indicate that the estimate, while problematic, is not necessarily enough to prevent the model being used. However, this is not the case here; the variance estimates are not quite large enough for the upper end of the CI to rise above 0, indicating a genuine problem, and this still leaves the impossibly large loading. The only solution found is to delete one of the two variables, as this eliminates the issue regardless of which one is removed. While total listing is causing the problems when both are used, it is a more informative measure and the preferred choice if only one is to be used.

Debt finance is initially measured in terms of Debt/Assets, Debt/Equity, and the total value of debt in the firm, not scaled by size. As anticipated from its unscaled nature, the third measure is problematic, in this case by providing almost no loading for the latent. It can be removed safely, which then leads to the same problem with Debt/Equity. Debt/Assets is the sole remaining measure of debt finance in the firm in most models.

Performance is initially measured using profit/size, earnings margin, and ROA. Generally, earnings margin demonstrates a tendency towards too-high loadings and negative variance and is removed for this reason. ROA is removed in some cases due to a low loading, but in others it has a higher effect and is retained.

Firm size retains five of its intended six indicators. With all six there is arguably a problem with discriminant validity here in that the Alpha for the measures chosen tends towards very high values, potentially indicating that two or more of the measures are too similar.

These findings are generally true across other models. Differences are noted where they arise.

Latent variable values and correlations with other variables were found and examined after each model was fitted. The predicted value of each latent variable was found to vary greatly with data type; a size latent using the base data is very different to one using the logarithmic data, for example. However, the values are almost identical between model types as long as the underlying data is the same; for example, the size latent variables in all four models below using the normal score data all have near-identical predicted values. The equations predicting latent variable values are therefore provided only for the exploratory models. The sole exception is a latent performance variable that was not part of the normal score exploratory model but was present in the normal score Signalling model. It is discussed at the model in which it appears.

By contrast, correlations involving latent variables are listed with each model. Two models using different data types often have different variables included due to different expectations regarding the underlying theories. Correlations are therefore more idiosyncratic to a single model.

SEM software offers an option known as modification indices (MI). Each MI is attached to a potential link between variables which was not included in the model and estimates the effects of including that link on the model's overall fit. The process calculates the expected decrease to the chi-square statistic if the model were to be re-fit with the new parameter included. This helps the researcher determine whether a connection between variables (either causal or covariance) should be included in later iterations of a model.

While useful, there are two matters to be aware of when using MIs. The first is that their use can render a model impossible to generalise. The use of MIs explains what will improve fit based entirely on the data used to form the model. Repeated use of MIs to iterate on a given model will ultimately lead to a final model that captures all fit improvements possible with the data in use and may suggest including links that only exist due to chance in the sample used to test the model. Taking the process too far may result in a model that almost perfectly describes the data used, but fits very poorly on another sample. The other potential problem is that MIs are calculated without regard to theory. MIs may suggest a linkage that improves the model fit but is not supported by theory. Some degree of judgement is required in determining whether a given MI should be accounted for in later iterations of a model.

In this thesis, MIs are used to search for model fit improvements. The value of the MI itself must be above (approximately) 3.84, the value corresponding to a chi-square test on one degree of freedom being significant at the 5% level. In addition, the expected parameter change must be substantial; some judgement is required to determine what is substantial for any given parameter, but many cases occurred where practically insignificant parameter changes (e.g. 0.01 or lower) were recorded. In addition, the connection suggested by the index must be between two variables that are valid variables to connect. For example, following the general rule that an indicator variable should be causally linked to only one latent variable, cases where this kind of connection are suggested are ignored. In addition, the suggested connection must be consistent with the theory that the model represents.

The initial form of the exploratory model series, named Latent in tables, is represented below:

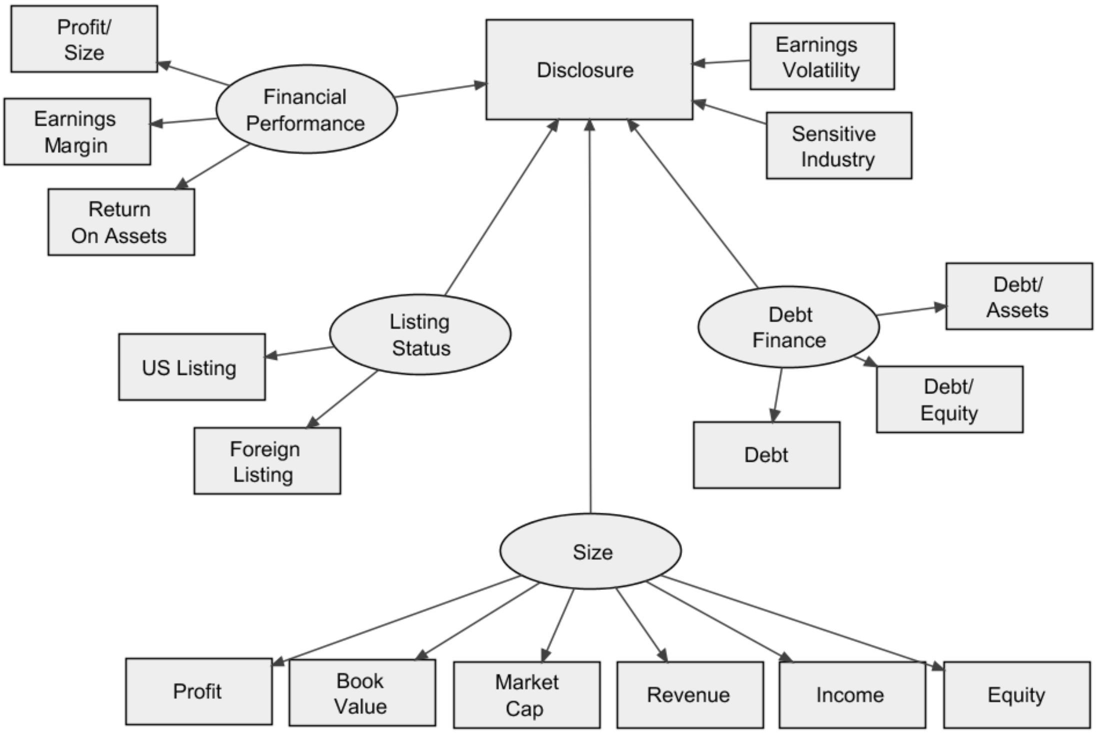

Figure 7.1: Latent Model Diagram, Initial Conception

7.2.1: Base Data

When modelled using the unaltered data, it quickly became apparent that the latent variables were generally not helping with measurement. Between the various reliability and validity tests above and the CFA itself, it soon emerged that size alone was benefiting from having multiple measurements.

The final SEM constructed is:

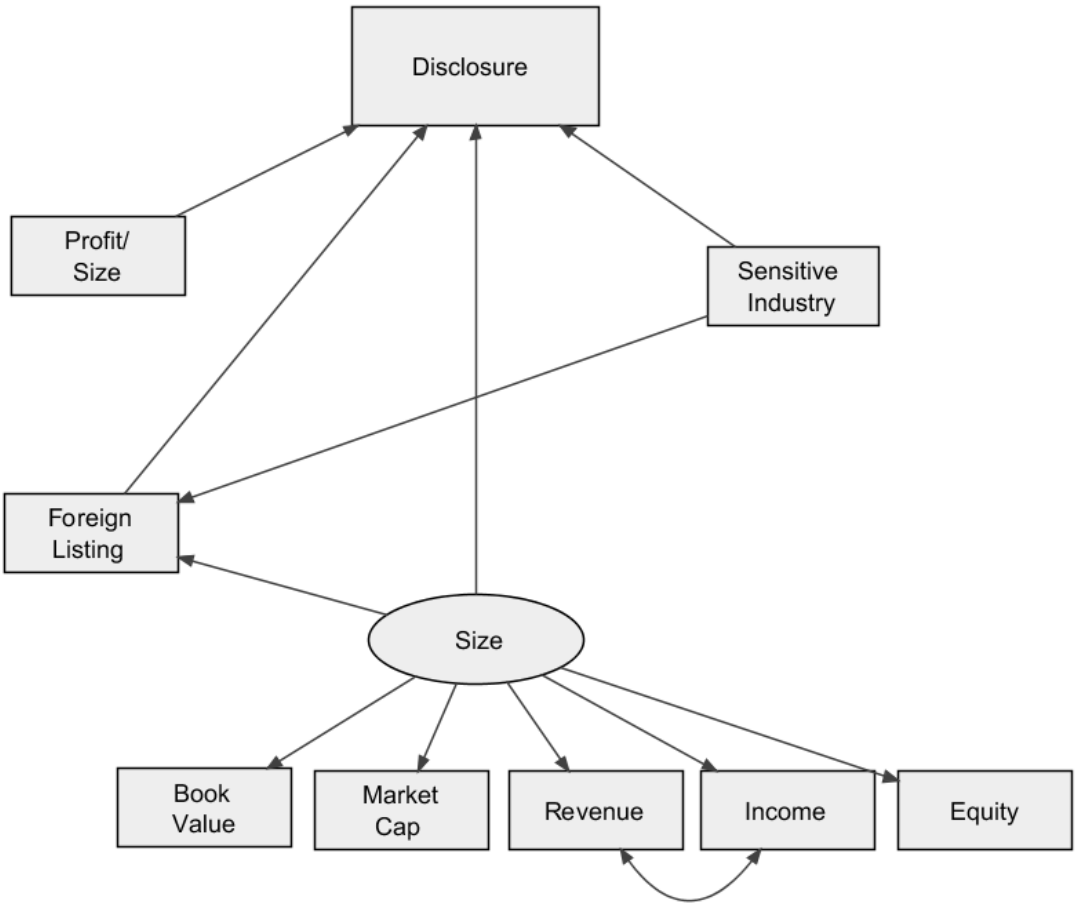

Figure 7.2: Latent Model, Final Version

Only the size latent remains in the end. This proves acceptable in most regards, with CR 0.912 (well above the required 0.7) and Average Variance Extracted 0.657. Cronbach's Alpha for this latent is 0.928, acceptable but high enough that there may be significant overlap in two or more of the indicators.

The additional connections in this model compared to figure 7.1 are the results of following modification indices. Allowing two indicators of a latent variable to covary often makes a large difference to the model's fit and is commonly seen in many models below. The other additions are allowing size and sensitive industry membership to cause foreign listing. Sensitivity is considered largely fixed for a given company; it is difficult to change whether a company is sensitive as this involves changing industry sector. Where a MI suggests allowing sensitivity to connect to another variable, sensitivity is always used as the causal variable. The link between size and foreign listing is not as clear, as discussed in chapter 5, but the MI suggestion was to allow size as the cause.

Table 7.2: Latent Model SEM estimates

			Estimate	S.E.	C.R.	P	Standardised Estimate
Foreign	<---	Sensitive	0.363	0.04	9.108	0.000	0.233
Foreign	<---	Size	0	0	3.905	0.000	0.100
Equity	<---	Size	1				0.974
Income	<---	Size	0.276	0.002	143.644	0.000	0.993
Revenue	<---	Size	0.747	0.015	48.229	0.000	0.801
Market Cap	<---	Size	2.407	0.016	148.53	0.000	0.995
Book Value	<---	Size	1.649	0.112	14.707	0.000	0.364
Following	<---	Sensitive	1.091	0.328	3.330	0.000	0.082
Following	<---	Size	0	0	4.826	0.000	0.116
Following	<---	Foreign	3.179	0.211	15.067	0.000	0.371
Following	<---	Profit/Size	0.302	0.108	2.794	0.005	0.067
Following	<---	Profit/Size	0.302	0.108	2.794	0.005	0.067

The only insignificant relationship left in the final model is the effect of volatility on analyst following. Looking at the standardised loadings, a few more pieces of information are clear. First, the size latent is dominated by income and market cap. Second, the size latent is not particularly powerful as a determinant of disclosure.

This model is intended to examine the potential latent variables rather than explore one of the theories of disclosure. It should not be expected to fit the data well.

Table 7.3: Latent Model Fit

Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
2177.565	24	0	90.732	0.841	0.764	0.561	0.25	0.005

As expected, the fit of this model is clearly lacking. The chi-square is high and leads to a hugely significant discrepancy, backed up by the chi/DF ratio being over 90. Neither set of indices demonstrates particularly good fit, falling a little short of the cut-off points. Most tellingly, however, is that the RMSEA is over double the maximum acceptable value. Further, the Bollen-Stine p-value is highly significant (down below the 1% level), indicating poor fit.

Table 7.1 shows that this model fits better than the regressions in terms of the index measures, but is notably worse in the RMSEA and Bollen-Stine p-value. Overall, the results suggest that there may be some usefulness to latent variables (at least for company size), but connections between variables may be necessary to obtain further improvements in model fit.

The only latent variable in any of the models using the unaltered data is a size latent and it does not vary much between models. Its predicted value is approximately:

Size = 0*Book Value + 0.233*Market Cap + 0.009*Revenue + 1.299*Income + 0.094*Equity (7)

In later models that include a Profit term, the coefficient attached to this is 0.005.

Outside of the size variables that define it, this variable correlates weakly with analyst following, foreign listing, and sensitive industry classification. Size does not correlate significantly with the remaining performance measure. This result is expected; any given latent variable should not correlated strongly with an observed variable that is not one of its indicators.

Pearson correlation coefficient values are provided below. The low correlation with profit suggests this is not a good indicator.

Table 7.4: Latent Correlations	
Latent	Size
Foreign listing	.102**
Sensitive	.130**
Book Value	.363**
Market Cap	.998**
Revenue	.801**
Income	.995**
Equity	.976**
Profit/Size	.012
Following	.155**

Although not reported in Table 7.1 above, another exploratory model was tested. This was visually similar to Figure 7.1 but added paths between latent variables where a correlation had previously been identified in Section 5.3. Taking the same approach of removing insignificant items from the model resulted in a final version that was nearly identical to Figure 7.2, differing only in that size was considered a cause of listing status. The resulting model was examined for estimates and fit, and the results had only trivial differences compared to Tables 7.2 and 7.3. For this reason, the results are not reported in Table 7.1.

7.2.2: Outliers Removed

When the same initial model is tested with the outliers removed, the final version is very different.

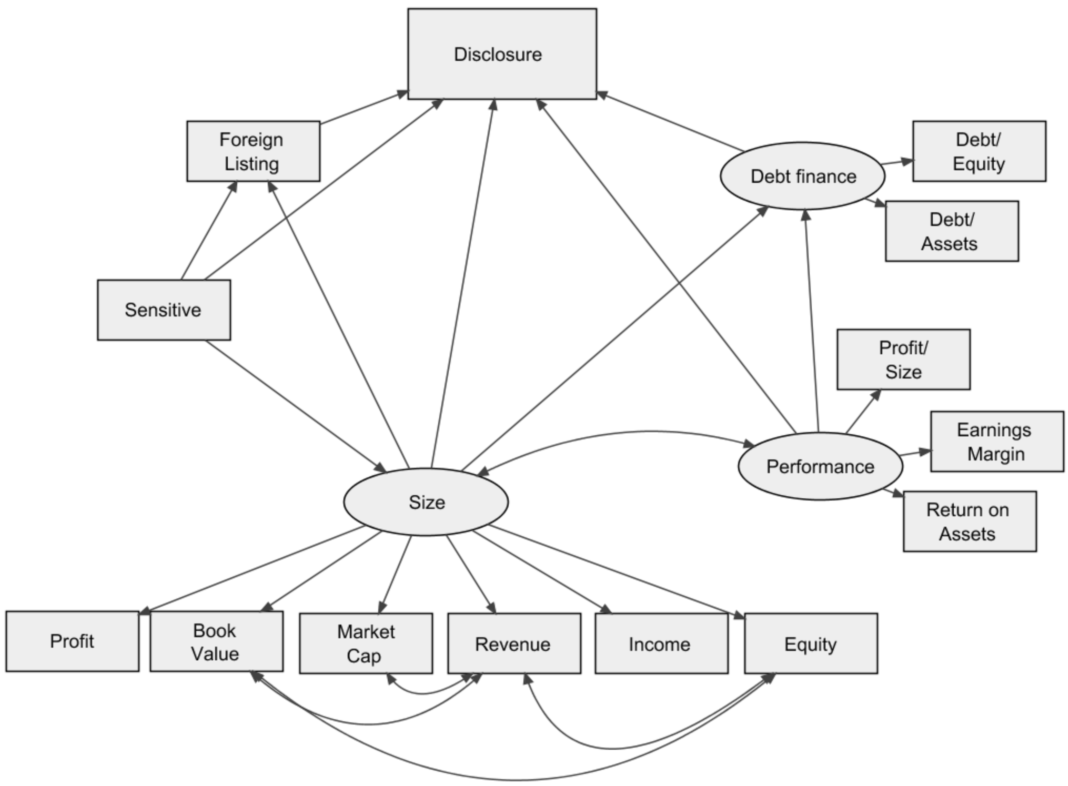

Figure 7.3: Final Model with Outliers Removed

Modification index examination of this model suggested many possible connections. As this model is based only on exploring the data and not any more direct disclosure theory, all are included. Correlations among indicators are not discussed. All are significant.

Company size is suggested to link to both debt finance and foreign listing. In both cases, size is assigned as the causal variable. Size is assumed to lead to or include foreign operations, which in turn encourage foreign listing. Debt finance is difficult to argue strongly in either direction; a company that has reached the size that equity alone will support may look to alternative sources of financing, but at the same time a company that has already secured debt finance has greater resources with which to grow. The connection can be argued to work in either direction; the MI suggests that size be considered the causal variable, however.

Sensitivity is suggested to cause foreign listing and financial performance. These two possible connections are accepted as suggested because the relative immutability of industrial sector compared to the other two variables means it is far more likely to be the cause than the effect in each case. This implies that the industries considered sensitive are generally more profitable and internationally-focused than other sectors.

Table 7.5: Latent Model SEM estimates, outliers removed							
			Estimate	S.E.	C.R.	P	Std. Estimate
Performance	<---	Sensitive	-0.036	0.014	-2.670	0.008	-0.074
Foreign	<---	Sensitive	0.290	0.040	7.287	0.000	0.19
Debt Finance	<---	Performance	0.343	0.144	2.383	0.017	0.12
Debt Finance	<---	Size	0.000	0.000	2.983	0.003	0.148
Foreign	<---	Size	0.000	0.000	7.089	0.000	0.188
Equity	<---	Size	1.000				0.76
Income	<---	Size	0.436	0.010	42.304	0.000	0.997
Revenue	<---	Size	1.687	0.044	38.604	0.000	0.763
Market Cap	<---	Size	3.517	0.089	39.296	0.000	0.932
Book Value	<---	Size	3.404	0.164	20.700	0.000	0.471
Profit	<---	Size	0.357	0.010	36.080	0.000	0.872
Return On Assets	<---	Performance	1.000				0.927
Earnings Margin	<---	Performance	1451.855	236.670	6.135	0.000	0.171
Profit/ Size	<---	Performance	1.232	0.053	23.421	0.000	0.932
Debt/ Equity	<---	Debt Finance	1.000				0.489
Debt/ Assets	<---	Debt Finance	0.128	0.036	3.567	0.000	0.404
Following	<---	Sensitive	0.656	0.290	2.264	0.024	0.052
Following	<---	Debt Finance	1.489	0.472	3.157	0.002	0.166
Following	<---	Performance	5.807	0.649	8.946	0.000	0.226
Following	<---	Size	0.000	0.000	11.585	0.000	0.296
Following	<---	Foreign	2.339	0.193	12.142	0.000	0.283
Performance	<->	Size	44363	9295	4.773	0.000	0.137

In this table, size is the most powerful cause of analyst following, but foreign listing is almost as influential.

Table 7.6: Latent Model Fit, outliers removed								
Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
1180.702	65	0	18.165	0.899	0.865	0.642	0.112	0.005

As table 7.6 demonstrates, this model is a better fit than the basic data version overall. The size latent for this model is described as:

Size = 0*Book Value + 0.009*Market Cap + 0.001*Revenue + 2.158*Income + 0.006*Equity (8)

Again, the coefficient of 0 with BV likely represents a miniscule value. However, in this case, the variable's value is dominated by the income measurement and all others have barely noticeable effects.

Comparing this latent to the two others, the Pearson correlation with debt finance is 0.122, a relatively high value, while the correlation with performance is a slightly lower 0.094. It otherwise has low correlations with all variables in the model except for its indicators and the disclosure measure, as shown in table 7.7 below. In addition, the correlation between this variable and the disclosure measure is relatively high for a non-indicator at 0.416.

The debt finance latent is:

Debt finance = 840390.4*Debt/Assets + 163076.9*Debt/Equity (9)

This variable shows a weak correlation of 0.077 with the performance latent discussed below and low correlations with all variables other than its indicators.

The performance latent is:

Performance = 0.469*Profit/Size + 0*Earnings Margin + 0.328*Return on Assets (10)

In addition to the latent correlations described in the preceding paragraphs, this has low correlations with all variables except for two of its three indicators. The correlation with earnings margin is low.

Table 7.7: Latent Correlations			
Latent	Size	Debt Finance	Perf.
Foreign listing	.211**	.016	.022
Sensitive	.128**	.028	-.053*
Profit	.874**	.091**	.126**
Book Value	.471**	.080**	.060*
Market Cap	.934**	.095**	.113**
Revenue	.758**	.119**	.125**
Income	.989**	.094**	.122**
Equity	.762**	.070**	.101**
Debt/Equity	.075**	.841**	.095**
Debt/Assets	.071**	.696**	.012
Profit/Size	.109**	.066*	.968**
Earnings Margin	.031	.074**	.163**
Return on Assets	.128**	.082**	.963**
Following	.416**	.148**	.278**
Size	1	.094**	.122**
Debt Finance	.094**	1	.077**
Perf.	.122**	.077**	1

7.2.3: Logarithmic Data

Retesting this model with logarithmic data results in the following:

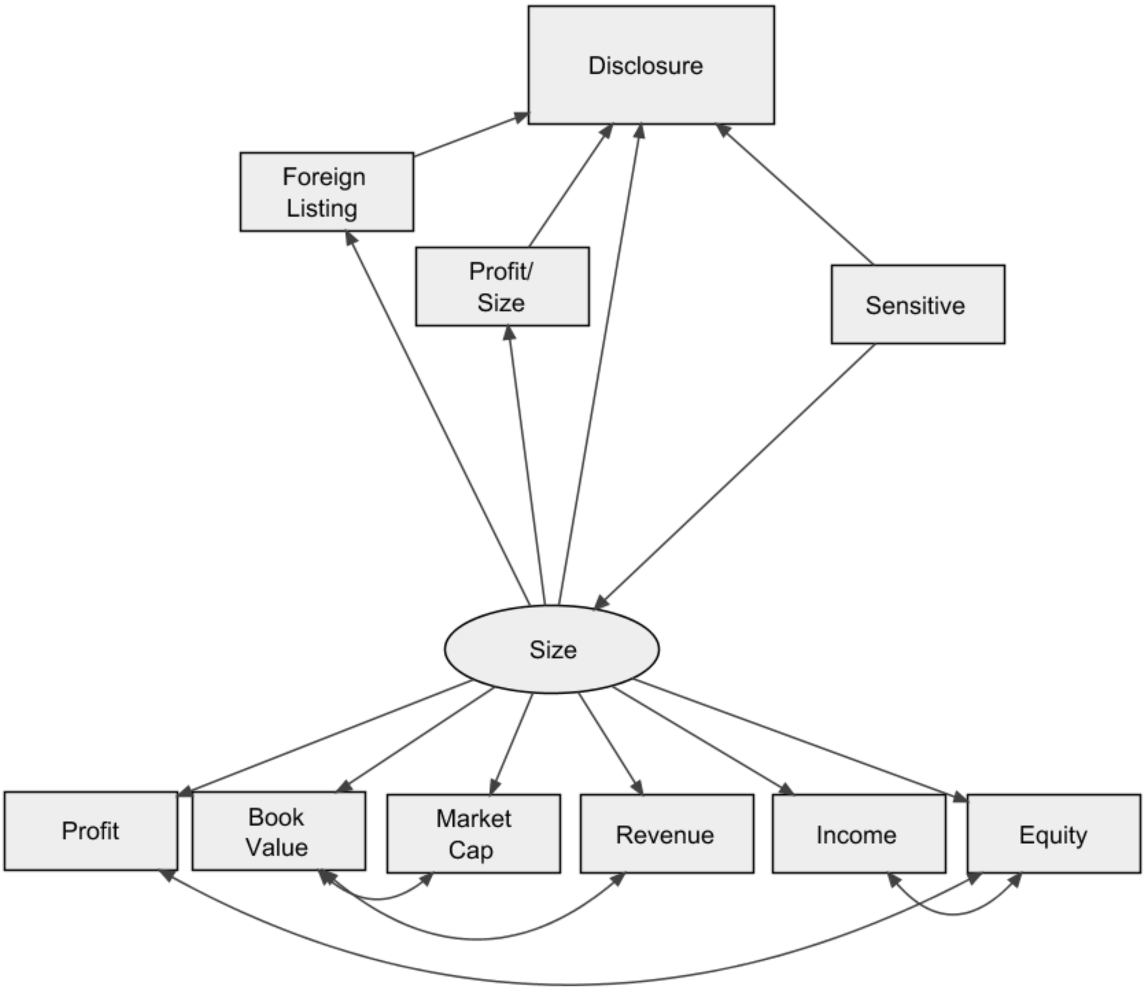

Table 7.4: Final Model With Logarithmic Data

MIs here suggest that size be linked to all other non-indicator variables. As no further theory defines the likely relationships, it is assumed that company size has a causal effect on all others. The sole exception is sensitivity. This variable represents the company's industry sector and is not easily changed. While it is possible that companies grow large through mergers that will push them into sensitive industries, it is far more likely that some of the industries considered sensitive here offer large growth potential. Sensitivity is therefore considered a cause of size.

The model's data and fit follows below.

Table 7.8: Latent Model SEM estimates, logarithmic data

			Estimate	S.E.	C.R.	P	Std. Estimate
Profit/ Size	<---	Size	0.002	0	4.985	0.000	0.135
Sensitive	<---	Size	0.102	0.019	5.516	0.000	0.15
Foreign	<---	Size	0.439	0.028	15.517	0.000	0.411
Equity	<---	Size	1				0.777
Income	<---	Size	0.247	0.023	10.945	0.000	0.293
Revenue	<---	Size	2.635	0.067	39.607	0.000	0.918
Market Cap	<---	Size	3.218	0.103	31.285	0.000	0.763
Book Value	<---	Size	3.388	0.084	40.271	0.000	0.93
Profit	<---	Size	0.14	0.01	13.374	0.000	0.355
Following	<---	Sensitive	0.43	0.208	2.063	0.039	0.032
Following	<---	Size	7.4	0.226	32.715	0.000	0.809
Following	<---	Foreign	0.543	0.146	3.721	0.000	0.063
Following	<---	Profit/ Size	-18.197	7.746	-2.349	0.019	-0.037

Table 7.9: Latent Model Fit, logarithmic data

Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
872.2	28	0	31.15	0.882	0.816	0.549	0.145	0.035

This has the overall second-best fit of the four versions of the exploratory model.

Only one latent variable remains in this model, that of size, and this is consistent across later logarithmic models. It is therefore impossible to compare this to other latent variables. It is described with the equation:

$$Size = 0.057*Log(Profit) + 0.098*Log(Book\ Value) + 0.023*Log(Market\ Cap) + 0.121*Log(Revenue) + 0.021*Log(Income) + 0.082*Log(Equity) \quad (11)$$

This latent variable is correlates to some extent with all variables in the model, as shown below. As with the previous exploratory models, correlations with non-indicators are low. Unlike the previous examples, two of the indicators have low correlations with the latent.

Table 7.10: Latent Correlations

Latent	Size
Foreign listing	.399**
Sensitive	.144**
Profit	.817**
Book Value	.945**
Market Cap	.788**
Revenue	.944**
Income	.379**
Equity	.308**
Profit/Size	-.178**
Following	-.069**

7.2.4: Normal Score Data

The final version of this model tested uses normal score data and results in the following:

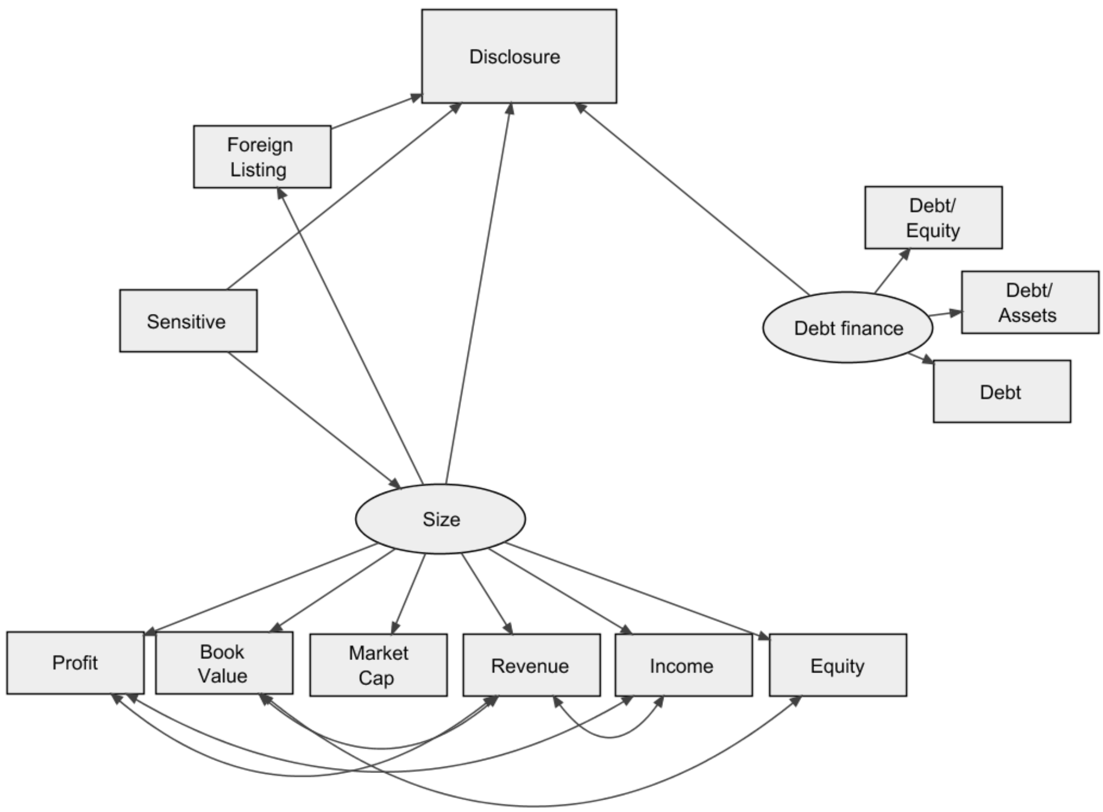

Table 7.5: Final Model with Normal Score Data

MIs suggest what is now clearly a pattern, allowing covariance among size indicators and adding the previously absent sensitivity-size link. No further MI-derived changes are present compared to the above models. The model's data and fit follows below.

Table 7.11: Latent Model SEM estimates, normal score data							
			Estimate	S.E.	C.R.	P	Std. Estimate
Size	<---	Sensitive	0.227	0.048	4.775	0.000	0.131
Foreign	<---	Size	0.375	0.024	15.916	0.000	0.416
Equity	<---	Size	1				0.865
Income	<---	Size	0.712	0.028	25.512	0.000	0.616
Revenue	<---	Size	0.899	0.024	36.849	0.000	0.797
Market Cap	<---	Size	1.027	0.023	43.945	0.000	0.896
Book Value	<---	Size	1.050	0.017	63.279	0.000	0.910
Profit	<---	Size	0.643	0.029	22.432	0.000	0.557
Debt	<---	Debt Finance	1				0.956
Debt/ Equity	<---	Debt Finance	0.792	0.027	29.602	0.000	0.720
Debt/ Assets	<---	Debt Finance	0.821	0.025	32.489	0.000	0.784
Following	<---	Sensitive	0.700	0.226	3.097	0.002	0.055
Following	<---	Size	5.246	0.171	30.64	0.000	0.717
Following	<---	Foreign	0.715	0.160	4.472	0.000	0.088
Following	<---	Debt Finance	0.897	0.132	6.789	0.000	0.125

Table 7.12: Latent Model Fit, normal score data								
Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
3122.814	46	0	67.887	0.791	0.703	0.551	0.216	0.050

This fit is not among the best for this model. The significant B-S p-value is unique to this version, but the other measures are all among the lowest for the exploratory model.

The size latent is here described with the equation:

Size = 0.02*Nscore(Profit) + 0.257*Nscore(Book Value) + 0.308*Nscore(Market Cap) + 0.049*Nscore(Revenue) + 0.034*Nscore(Income) + 0.137*Nscore(Equity) (12)

While the debt finance latent is:

Debt finance = 0.137*Nscore(Debt/Assets) + 0.095*NScore(Debt/Equity) + 0.74*NScore(Debt) (13)

The size latent correlates to a low extent with most of the variables in the model as shown below. However, the correlation between the value of debt and the size latent is a high 0.700, close to the level at which the size latent correlates to its own indicators. This provides further evidence that the value of debt is providing an indication of size. As a result, the correlation between the size and debt latent variables is also high at 0.629. The value of debt is not used as an indicator of debt finance in any further models.

The debt finance latent demonstrates further problems. It correlates strongly with some of the size indicators, particularly book value and revenue. This is another result of the debt value variable acting as a size measure.

Table 7.13: Latent Correlations		
Latent	Size	Debt Finance
Foreign listing	.393**	.202**
Sensitive	.116**	.063*
Profit	.576**	.383**
Book Value	.961**	.673**
Market Cap	.932**	.490**
Revenue	.837**	.684**
Income	.633**	.483**
Equity	.902**	.525**
Debt	.700**	.987**
Debt/Equity	.411**	.758**
Debt/Assets	.216**	.825**
Following	.461**	.532**
Size	1	.629**
Debt	.629**	1

7.3: Signalling Theory Models

The initial model of Signalling theory is as pictured below:

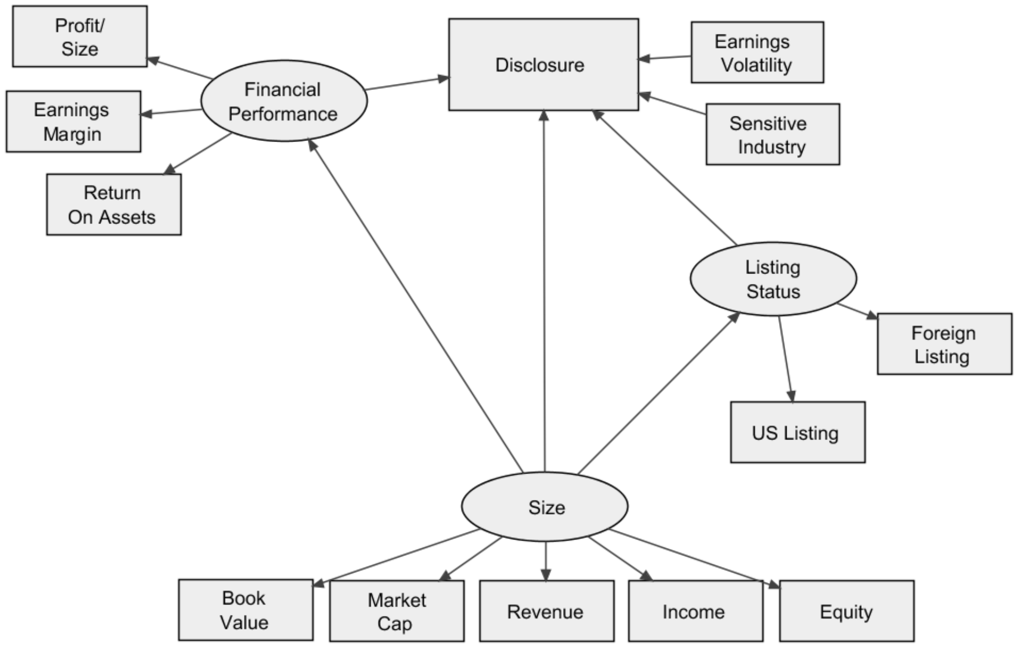

Figure 7.6: Signalling Theory, Initial Model

The basic idea behind this model is, as the theory itself suggests, a company with good news of any form will share it. Within the limitations of the model, the main indication of good news comes from the performance latent as this directly demonstrates the financial results.

Size is included because it may offer some good news in terms of diversification, offering the company stability and the potential to take losses in one area without the whole entity suffering. Listing status is included for similar reasons, in this case looking only at geographic diversification. Sensitivity has been included on the assumption that a more sensitive company may find it has more to gain from signalling. Volatility is included as investors like to see stable returns, so the ability to report low volatility is good news for the firm.

As shown above, size is considered a cause of listing and performance. Listing is easily explained in that a larger company has greater potential to have the resources needed to expand into a foreign market. Alternatively, a large company is more likely to attract international attention and therefore list elsewhere even if operations are limited to one location. The arrow could be argued to point in the opposite direction or even be a two-way link (see Appendix A), however; listing in a second market offers greater potential for funding that can be used to grow the company.

The size-performance link has a different justification. Size is assumed to potentially indicate lines of business and, from that, diversification. Like the justification for size as diversity, a failure in one part of the business does not harm the overall performance much.

7.3.1: Base Data

After investigation, the final model is:

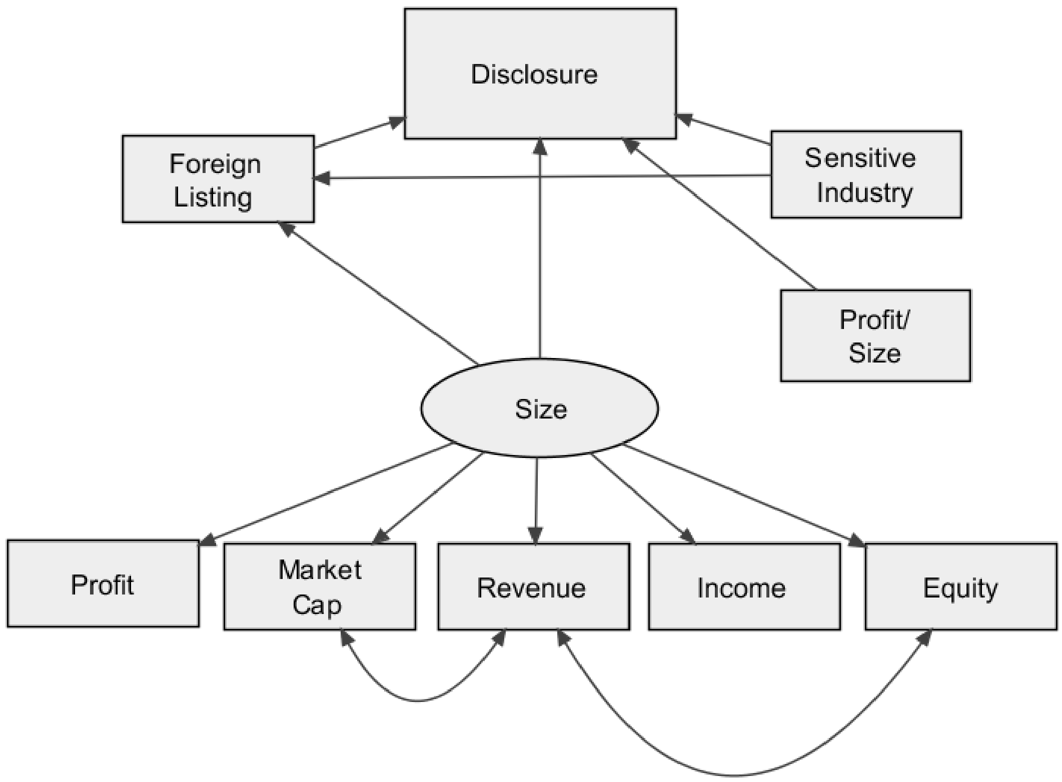

Figure 7.7: Signalling Theory, Final Model

The main MI-derived change here is allowing sensitivity to cause foreign listing. Almost everything has remained in the model, with the exceptions being a few indicators and the size-performance link. Book Value of Assets has been removed from the model and replaced with Profit between the initial model and this version.

Experimentation showed that Book Value contributed very little to this model, while Profit has a significant effect on the size latent variable.

Table 7.14: Signalling Theory Estimates

			Estimate	S.E.	C.R.	P	Std. Estimate
Foreign	<---	Size	0.000	0.000	3.88	0.000	0.099
Foreign	<---	Sensitive	0.363	0.04	9.109	0.000	0.233
Equity	<---	Size	1				0.973
Income	<---	Size	0.277	0.002	143.678	0.000	0.993
Revenue	<---	Size	0.747	0.016	48.201	0.000	0.801
Market Cap	<---	Size	2.407	0.016	146.847	0.000	0.995
Profit	<---	Size	0.297	0.014	21.215	0.000	0.493
Following	<---	Sensitive	1.091	0.328	3.331	0.000	0.082
Following	<---	Size	0	0	4.799	0.000	0.115
Following	<---	Foreign	3.18	0.211	15.073	0.000	0.371
Following	<---	Profit/ Size	0.302	0.108	2.795	0.005	0.067

Once again, a single item not significant at 5% has remained in the model and it is the volatility-disclosure link. Given that volatility was never significant in the regressions, this is not surprising. The effect of size on foreign listing is high enough that it cannot be dismissed, but is somewhat low nonetheless. In addition, note that sensitivity has a small effect.

In terms of the theory, there is an important finding. Performance is the second most powerful explanation of disclosure once standardisation is taken into account. This lends some weight to the theory, although the greater power of the Foreign Listing variable suggests another explanation of disclosure may be more valid.

Table 7.15: Signalling Theory Fit								
Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
180.239	23	0	7.836	0.986	0.98	0.631	0.069	0.562

According to the index measures, this model fit is excellent. RMSEA and the chi-square derived measures are still too high, although they are better than the Latent model. The B-S p-value, on the other hand, indicates a low but significant good fit.

The size latent variable is slightly different to the size latent in equation (7) in section 7.2.1, replacing the book value term with 0.005*Profit. The size latent variable's correlations are shown below. As usual, it correlates weakly with most variables and strongly with its indicators. The correlation with profit is weak for an indicator, however.

Table 7.16: Latent Correlations	
Latent	Size
Foreign listing	.104**
Sensitive	0.14
Profit	.493**
Market Cap	.804**
Revenue	.995**
Income	.974**
Equity	0.012
Profit/Size	.158**
Following	.104**

7.3.2: Outliers Removed

If modelled with the outliers removed, the model makes one important change: the performance variable is replaced with its latent:

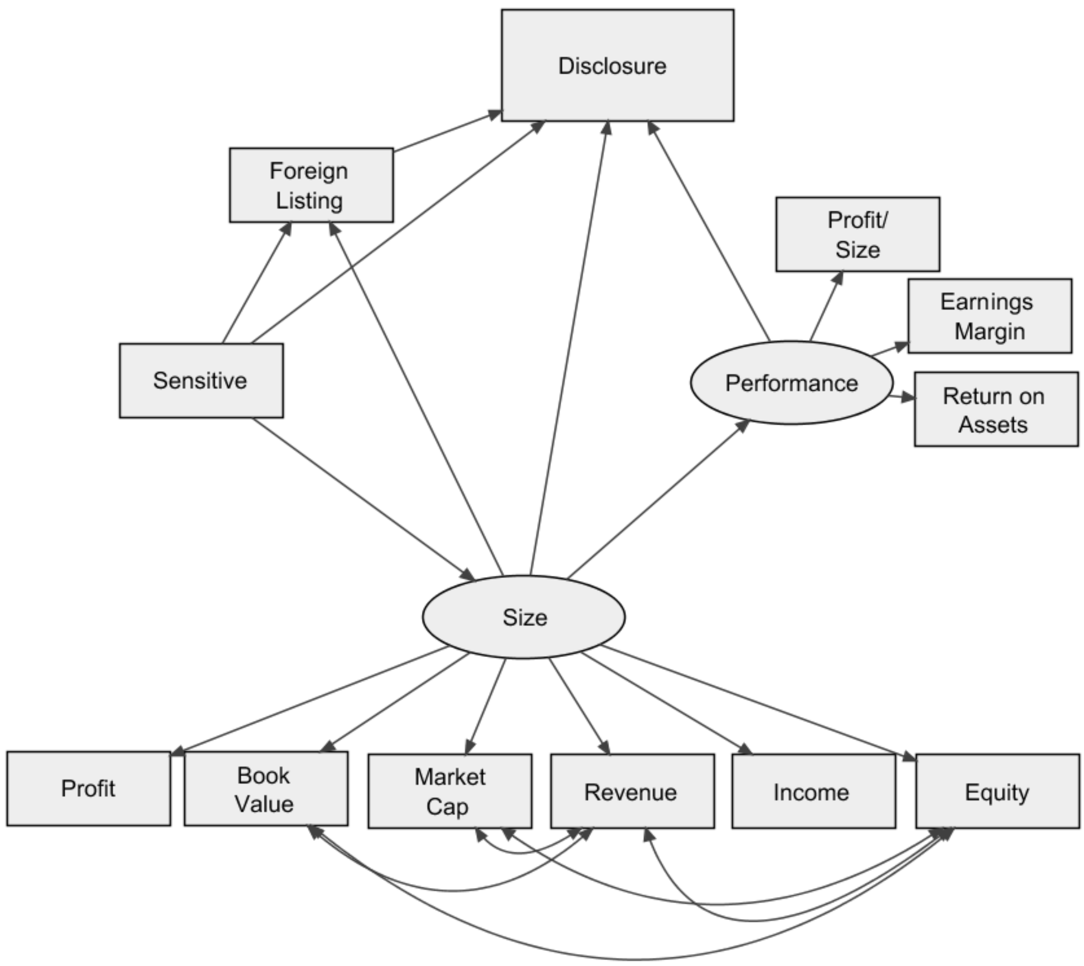

Figure 7.8: Final Model With Outliers Removed

No new MI-derived changes are made. However, this seemingly small change makes some important differences between this and the base data model.

Table 7.17: Signalling Model SEM estimates, outliers removed

			Estimate	S.E.	C.R.	P	Std. Estimate
Size	<---	Sensitive	339478.3	71315.26	4.760	0.000	0.128
Performance	<---	Size	0.000	0.000	4.450	0.000	0.125
Foreign	<---	Size	0.000	0.000	6.946	0.000	0.183
Foreign	<---	Sensitive	0.290	0.040	7.245	0.000	0.189
Equity	<---	Size	1				0.759
Income	<---	Size	0.440	0.010	42.326	0.000	1.004
Revenue	<---	Size	1.677	0.042	40.074	0.000	0.751
Market Cap	<---	Size	3.495	0.076	46.014	0.000	0.924
Book Value	<---	Size	3.404	0.163	20.825	0.000	0.470
Profit	<---	Size	0.356	0.010	36.059	0.000	0.868
Return on Assets	<---	Performance	1				0.925
Earnings Margin	<---	Performance	1452.095	237.460	6.115	0.000	0.171
Profit/ Size	<---	Performance	1.237	0.055	22.651	0.000	0.934
Following	<---	Sensitive	0.682	0.294	2.318	0.020	0.054
Following	<---	Performance	6.389	0.627	10.189	0.000	0.246
Following	<---	Size	0.000	0.000	12.567	0.000	0.305
Following	<---	Foreign	2.364	0.195	12.147	0.000	0.285

The performance latent in this model has a far more powerful effect on disclosure than the single variable in the base data model, suggesting the theory is more powerful if a latent variable can be involved. Other variables have stronger effects, but the difference is not large enough to suggest the performance latent is weak.

Table 7.18: Signalling Model Fit, outliers removed								
Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
915.613	44	0	20.809	0.921	0.886	0.614	0.120	0.005

Despite the explanatory power, this model does not fit as well as the base data model.

The size latent for this model is as equation (7) with the addition of a 0.005*Profit term. All further size latent variables follow this pattern.

Correlations for the two latent variables are provided below. As usual, neither correlates strongly with any observed variable other than its own indicators. Each latent also has a low correlation with one of its indicators.

Table 7.19: Latent Correlations		
Latent	Size	Perf.
Foreign listing	.208**	.021
Sensitive	.128**	-.053*
Profit	.864**	.126**
Book Value	.466**	.060*
Market Cap	.920**	.113**
Revenue	.748**	.125**
Income	.990**	.121**
Equity	.755**	.101**
Profit/Size	.107**	.970**
Earnings Margin	.032	.164**
Return on Assets	.127**	.961**
Following	.404**	.277**
Size	1	.120**
Performance	.120**	1

7.3.3: Logarithmic Data

Logarithmic data produces a model similar to the base data model:

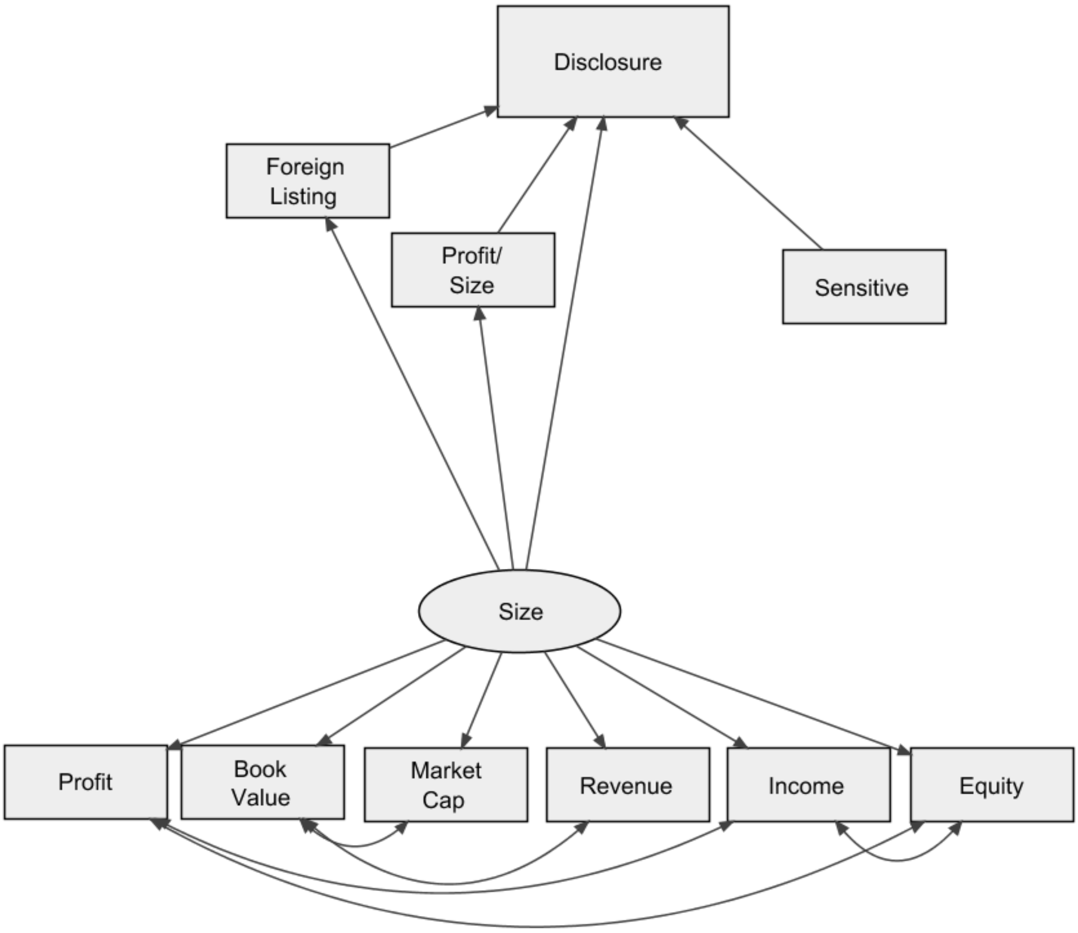

Figure 7.9: Final Model with Logarithmic Data

No MI-derived changes beyond size covariances are included. The difference is size having a causal effect on the profit/size variable in this model.

Table 7.20: Signalling Model SEM estimates, logarithmic data

			Estimate	S.E.	C.R.	P	Std. Estimate
Foreign	<---	Size	0.445	0.028	15.627	0.000	0.425
Profit/ Size	<---	Size	0.002	0.001	4.427	0.000	0.124
Equity	<---	Size	1				0.792
Income	<---	Size	0.243	0.022	10.961	0.000	0.294
Revenue	<---	Size	2.497	0.067	37.158	0.000	0.886
Market Cap	<---	Size	3.08	0.102	30.085	0.000	0.745
Book Value	<---	Size	3.116	0.086	36.264	0.000	0.872
Profit	<---	Size	0.135	0.010	13.515	0.000	0.348
Following	<---	Sensitive	0.477	0.200	2.384	0.017	0.036
Following	<---	Size	7.716	0.226	34.176	0.000	0.864
Following	<---	Foreign	0.255	0.149	1.708	0.088	0.03
Following	<---	Profit/ Size	-16.509	7.658	-2.156	0.031	-0.033

As in the base data model, the single performance variable has a weak effect on disclosure. In this case it is negative.

Table 7.21: Signalling Model Fit, logarithmic data

Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
312.755	28	0.000	11.170	0.958	0.938	0.596	0.084	0.005

In absolute terms, this model fits well. The ratio measures are high and the RMSEA only marginally misses the 0.08 required minimum. However, the base data fits better.

The size latent variable for this model shows the expected weak correlations with non-indicators. As in section 7.2.3, the correlation between the latent and two of the indicators is low.

Table 7.22: Latent Correlations	
Latent	Size
Foreign listing	.399**
Sensitive	.145**
Profit	.819**
Book Value	.937**
Market Cap	.791**
Revenue	.943**
Income	.371**
Equity	.307**
Profit/Size	-.175**
Following	-.068*

7.3.4: Normal Score Data

Finally, using normal score data results in a model similar to the outlier-removed model:

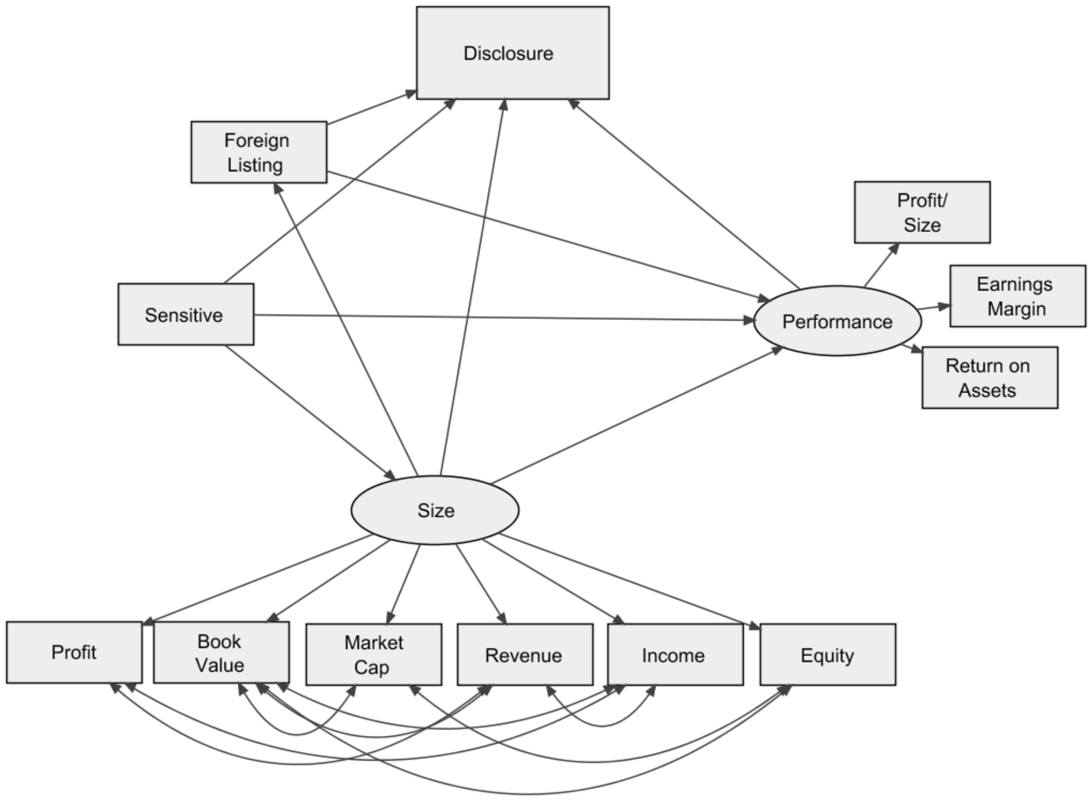

Figure 7.10: Final Model with Normal Score Data

In a rare example, MIs on the initial version of this model suggest adding some unusual items. As normal, a size-sensitivity link and size indicator covariances are suggested. In addition, a number of modification indices relate to the performance latent. Indices suggest that all three of the other explanatory variables (size, sensitivity, foreign listing) may have some connection to the performance latent.

The suggestion given is that each of these variables causes performance. Sensitivity is relatively immutable as it is defined by industry sector, so this variable is allowed as a cause of performance. The remaining two links are less clear, however. A larger company may be able to benefit from economies of scale and have the efficiencies these create translate into higher income per unit of size. On the other hand, higher performance may lead to greater growth potential. The former explanation is chosen for use in the model.

Similarly, the listing-performance link can be argued to act in either direction. The assumption throughout the thesis is that foreign listing is strongly correlated with foreign operations, which in turn indicate the company's overall size. Like the direct size link argued above, it is possible that high performance leads to the resources required to expand into a new market. Alternatively, the company may expand into a foreign market because analysis suggests greater profit potential by doing so. Like the size argument above, the decision was made to allow listing as a cause of performance.

The addition of these MI-derived changes to the structural model has some implications for the theory in this case. Financial performance, the main driver of disclosure under the theory, is itself heavily influenced by other company characteristics. The Signalling part of disclosure is not the entire story, but the result of large quantities of company information

Table 7.23: Signalling Model SEM estimates, Normal Score Data

			Estimate	S.E.	C.R.	P	Std. Estimate
Size	<---	Sensitive	0.193	0.048	4.051	0.000	0.109
Foreign	<---	Size	0.337	0.023	14.753	0.000	0.380
Performance	<---	Size	0.64	0.028	22.523	0.000	0.627
Performance	<---	Foreign	-0.226	0.028	-8.119	0.000	-0.196
Performance	<---	Sensitive	-0.208	0.04	-5.181	0.000	-0.116
Equity	<---	Size	1				0.878
Income	<---	Size	0.784	0.025	31.077	0.000	0.689
Revenue	<---	Size	0.966	0.021	46.15	0.000	0.866
Market Cap	<---	Size	0.963	0.022	44.655	0.000	0.852
Book Value	<---	Size	1.070	0.019	55.148	0.000	0.940
Profit	<---	Size	0.714	0.026	27.197	0.000	0.627
Return on Assets	<---	Performance	1				0.899
Earnings	<---	Performance	0.865	0.022	39.668	0.000	0.777
Profit/ Size	<---	Performance	1.101	0.018	61.551	0.000	0.990
Following	<---	Sensitive	0.724	0.227	3.193	0.001	0.054
Following	<---	Performance	-0.551	0.158	-3.495	0.000	-0.074
Following	<---	Size	5.945	0.196	30.393	0.000	0.784
Following	<---	Foreign	0.806	0.160	5.041	0.000	0.094

The inclusion of a performance latent variable has in this case not led to a strong effect, with it having a low standardised estimate of the effect of performance on analyst following.

Table 7.24: Signalling Model Fit, Normal Score Data								
Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
1661.922	41	0.000	40.535	0.894	0.832	0.555	0.166	0.005

The fit of this model is low. Few of the models tested have NFI or TLI below 0.9 and the RMSEA of this model is double the maximum acceptable value.

Uniquely, this model contains a latent variable that was not present in the comparable exploration model, that of performance. It is described by the equation:

$$\text{Performance} = 0.398*\text{NScore}(\text{Profit/Size}) + 0.428*\text{NScore}(\text{ROA}) \quad (14)$$

The coefficient for Earnings Margin is 0. While such variables were included in some earlier models, in this case the rescaling performed as part of the Normal Score transformation means the coefficient is not related to the large scale of the variable.

Correlations involving the latent variables in this model demonstrate that each latent correlates strongly to its indicators. However, in this model the correlations with other variables are stronger than usual and close to being problematically high.

Table 7.25: Latent Correlations		
Latent	Size	Perf.
Foreign listing	.290**	0.016
Sensitive	.055*	-.095**
Profit	.765**	.466**
Book Value	.883**	.452**
Market Cap	.835**	.308**
Revenue	.932**	.444**
Income	.822**	.436**
Equity	.814**	.292**
Profit/Size	.400**	.997**
Earnings Margin	.501**	.791**
Return on Assets	.540**	.921**
Following	.473**	.346**
Size	1	.421**
Performance	.421**	1

7.4: Agency Theory

The initial Agency model is:

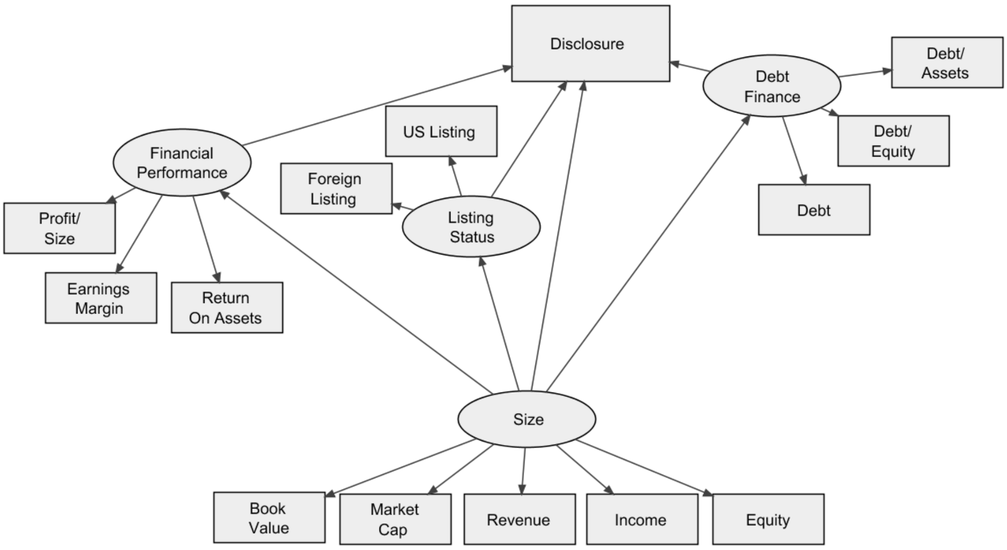

Figure 7.11: Agency Theory, Initial Model

This model is based on the idea that anything that may cause agency problems within the firm is a cause of disclosure. Other potential causes are not necessarily untrue, but are not explained by Agency theory. Debt finance and listing status are the most important items here. Debt is directly used in the original explanation of the theory, so in disclosure terms is a clear indication of an Agency problem situation that requires disclosure of activity to counteract. Listing status is included because the more owners there are, the less interest each has in making sure the company is working to their benefit at all times. Performance is included because of the argument that good performance needs no explanation. If this is true, good results would imply lower vigilance among owners, potentially enabling more self-interest among managers.

Size is included, but only as a cause of the other three. There is the possibility that size causes agency simply because a larger firm is more difficult to monitor completely, but this is not taken into account in the model as it exacerbates agency problems without causing them. The links between size and performance and size and listing are similar to those explained under Signalling above.

7.4.1: Base Data

The model is quickly shown to be very inaccurate and is greatly cut down:

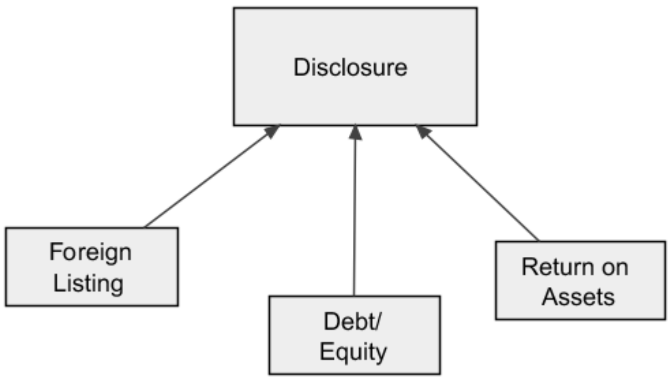

Figure 7.12: Agency Theory, Final Model

No significant or large MI-derived alterations were suggested. All of the remaining variables in this model are single-indicators in place of latent variables.

Table 7.26: Agency Theory Estimates

	Estimate	S.E.	C.R.	P	Std. Estimate
Following <--- Foreign	3.384	0.205	16.545	0.000	0.396
Following <--- Return on Assets	1.486	0.252	5.900	0.000	0.141
Following <--- Debt/Equity	0.003	0.029	0.102	0.919	0.002

With the model as stripped back as it is, drawing conclusions becomes difficult. However, based on the information given, it appears Agency is not supported as an explanation of disclosure. The measure of debt finance usage, Debt/Equity, is both very weak as a standardised estimate and very insignificant. The lack of debt finance as an explanation of disclosure suggests the theory is not true in this context.

The model does not contain any size measure as this became insignificant early in the modelling process. This is very unusual and goes against the near-universal finding in literature that size is a positive and significant determinant of disclosure.

Table 7.27: Agency Theory Fit								
Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
4.797	3	0.187	1.599	0.984	0.987	0.492	0.020	0.069

Perhaps because there is so little left in the model, the fit here is excellent, being by some margin the best model. Almost all tests suggest this model to be acceptable and very well-fitting despite being largely meaningless as a model of Agency Theory. The B-S p-value is significant but oddly low, possibly indicating that the non-normality of the data is inflating the fit in other measures.

However the PNFI is below 0.5. The reason for this is unclear as it is not consistent with any other fit measure. A low PNFI for an otherwise well-fitting but highly complex model would be expected, but this is the simplest model examined in the thesis.

No latent variable correlations are discussed as no latent variable was present in the final version of this model.

7.4.2: Outliers Removed

Removing outliers results in a very different model.

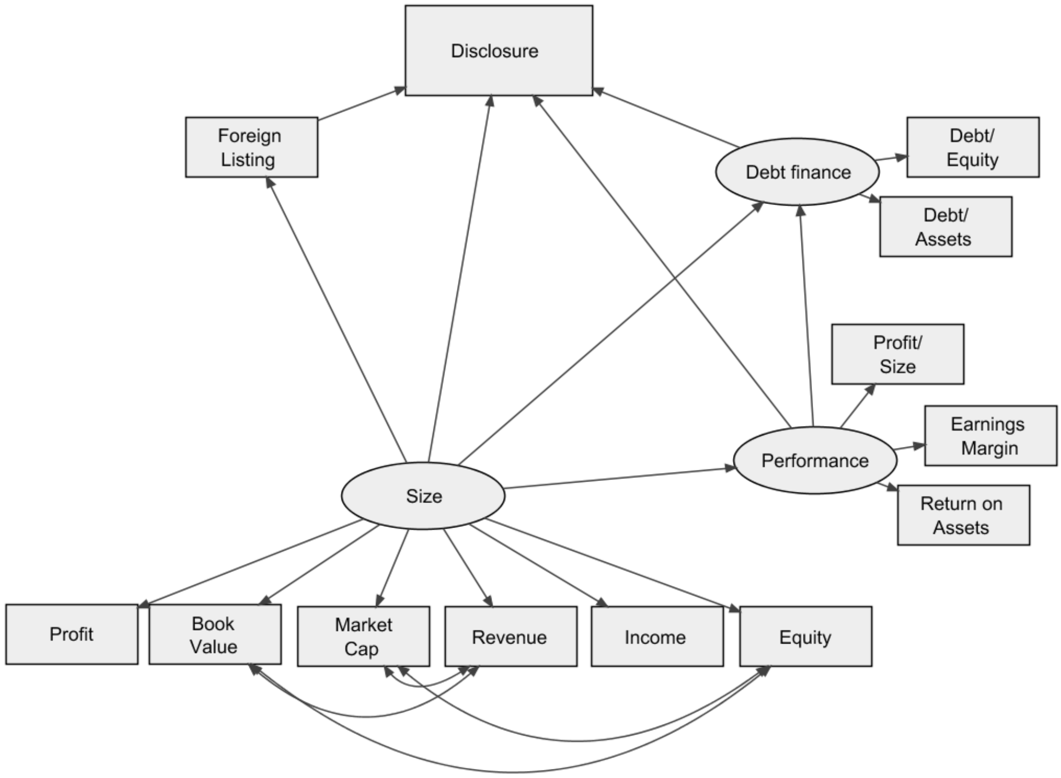

Figure 7.13: Final Model with Outliers Removed

This is one of the more complete models in terms of how little has been removed compared to the original conception of the model.

MIs here suggest the usual covariances among size indicators, plus the unusual case of allowing performance as a cause of debt finance. The connection makes some logical sense; a higher-performing company will be assumed to be more capable of meeting loan interest payments and may therefore be able to acquire more credit and more favourable repayment terms. Higher performance could be taken as a sign of lower default risk, reducing the cost of this form of capital.

Table 7.28: Agency Model SEM estimates, outliers removed

			Estimate	S.E.	C.R.	P	Std. Estimate
Performance	<---	Size	0.000	0.000	4.458	0.000	0.125
Debt Finance	<---	Size	0.000	0.000	2.94	0.003	0.144
Foreign	<---	Size	0.000	0.000	7.758	0.000	0.208
Debt Finance	<---	Performance	0.347	0.144	2.409	0.016	0.121
Equity	<---	Size	1				0.759
Income	<---	Size	0.440	0.010	42.305	0.000	1.004
Revenue	<---	Size	1.677	0.042	40.069	0.000	0.751
Market Cap	<---	Size	3.496	0.076	46.004	0.000	0.925
Book Value	<---	Size	3.404	0.163	20.823	0.000	0.470
Profit	<---	Size	0.356	0.010	36.055	0.000	0.869
Return on Assets	<---	Performance	1				0.927
Earnings Margin	<---	Performance	1446.708	237.25	6.098	0.000	0.170
Profit/ Size	<---	Performance	1.233	0.055	22.527	0.000	0.933
Debt Equity	<---	Debt Finance	1				0.490
Debt Assets	<---	Debt Finance	0.127	0.036	3.572	0.000	0.403
Following	<---	Debt Finance	1.527	0.478	3.196	0.001	0.170
Following	<---	Performance	5.752	0.653	8.807	0.000	0.222
Following	<---	Size	0.000	0.000	11.262	0.000	0.286
Following	<---	Foreign	2.452	0.19	12.898	0.000	0.296

Overall, this model implies that debt financing does have an effect on disclosure, but that debt is influenced by other factors. Agency Theory has some explanatory power, but there are other forces that influence the extent to which agency costs affect the company. The debt finance latent is the weakest of those that influence analyst following, however. The higher power of the financial performance latent suggests that Signalling Theory may better explain disclosure, but that both theories have some power and disclosure decisions may be the result of both together.

Table 7.29: Agency Model Fit, outliers removed

Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
918.821	54	0	17.015	0.885	0.891	0.637	0.108	0.010

The fit for this model is generally lower than that of the base data, but as this model contains more variables and includes debt finance, it provides more useful information on the theory.

All three of the latent variables in this model correlate strongly with their indicators and weakly or not at all with all other variables. As in other outlier-free data models, the correlation between size and analyst following is high for a non-indicator variable.

Table 7.30: Latent Correlations			
Latent	Size	Debt Finance	Perf.
Foreign listing	.211**	.016	.022
Sensitive	.128**	.028	-.053*
Profit	.874**	.091**	.126**
Book Value	.471**	.080**	.060*
Market Cap	.934**	.095**	.113**
Revenue	.758**	.119**	.125**
Income	.989**	.094**	.122**
Equity	.762**	.070**	.101**
Debt/Equity	.075**	.841**	.095**
Debt/Assets	.071**	.696**	.012
Profit/Size	.109**	.066*	.968**
Earnings Margin	.031	.074**	.163**
Return on Assets	.128**	.082**	.963**
Following	.416**	.148**	.278**
Size	1	.094**	.122**
Debt Finance	.094**	1	.077**
Perf.	.122**	.077**	1

7.4.3: Logarithmic Data

Logarithmic data provides a different view of the theory:

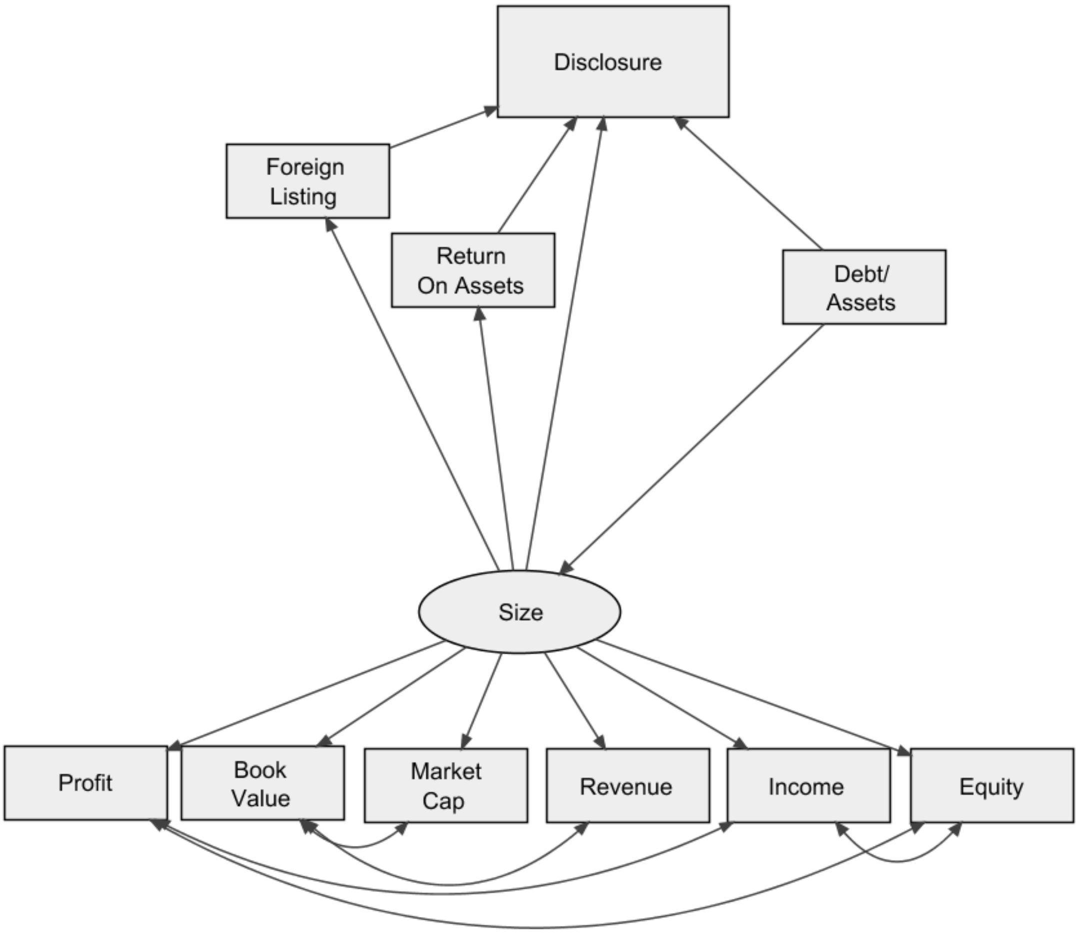

Figure 7.14: Final Model with Logarithmic Data

This model is reduced compared to that using untransformed but outlier-free data, but contains far more than the base data model. Importantly, a debt measure is present, so this model may provide information about the theory. No MI-derived changes other than covariances within size were significant.

Table 7.31: Agency Model SEM estimates, logarithmic data

			Estimate	S.E.	C.R.	P	Std. Estimate
Foreign	<---	Size	0.446	0.029	15.53	0.000	0.421
Return on Assets	<---	Size	0.007	0.001	9.823	0.000	0.271
Debt/ Assets	<---	Size	-0.455	0.067	-6.834	0.000	-0.19
Equity	<---	Size	1				0.784
Income	<---	Size	0.241	0.022	10.777	0.000	0.288
Revenue	<---	Size	2.534	0.069	36.95	0.000	0.89
Market Cap	<---	Size	3.13	0.104	30.072	0.000	0.749
Book Value	<---	Size	3.202	0.087	36.626	0.000	0.886
Profit	<---	Size	0.134	0.01	13.377	0.000	0.343
Following	<---	Size	8.035	0.24	33.434	0.000	0.887
Following	<---	Foreign	0.305	0.149	2.05	0.04	0.036
Following	<---	Return on Assets	-28.789	5.696	-5.054	0.000	-0.081
Following	<---	Debt/ Assets	0.222	0.059	3.766	0.000	0.059

While size is ultimately far more important, as indicated by its large standardised estimate, the debt/assets ratio is a determinant of analyst following.

Table 7.33: Agency Model Fit, logarithmic data

Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
348.056	27	0	12.891	0.954	0.929	0.572	0.091	0.050

This fit table is, on the whole, the second best of those for Agency Theory models. The base data does have better fit, but provides little information about the theory. This model is therefore superior.

Correlations for the size latent are presented below. As in other logarithmic data models, the correlations with two of the indicators are low.

Table 7.34: Latent Correlations	
Latent	Size
Foreign listing	.400**
Profit	.817**
Book Value	.946**
Market Cap	.789**
Revenue	.945**
Income	.365**
Equity	.299**
Debt/Assets	0.004
Return on Assets	-0.007
Following	-.068**

7.4.4: Normal Score Data

Normal score data for this theory results in the same problem as the initial model:

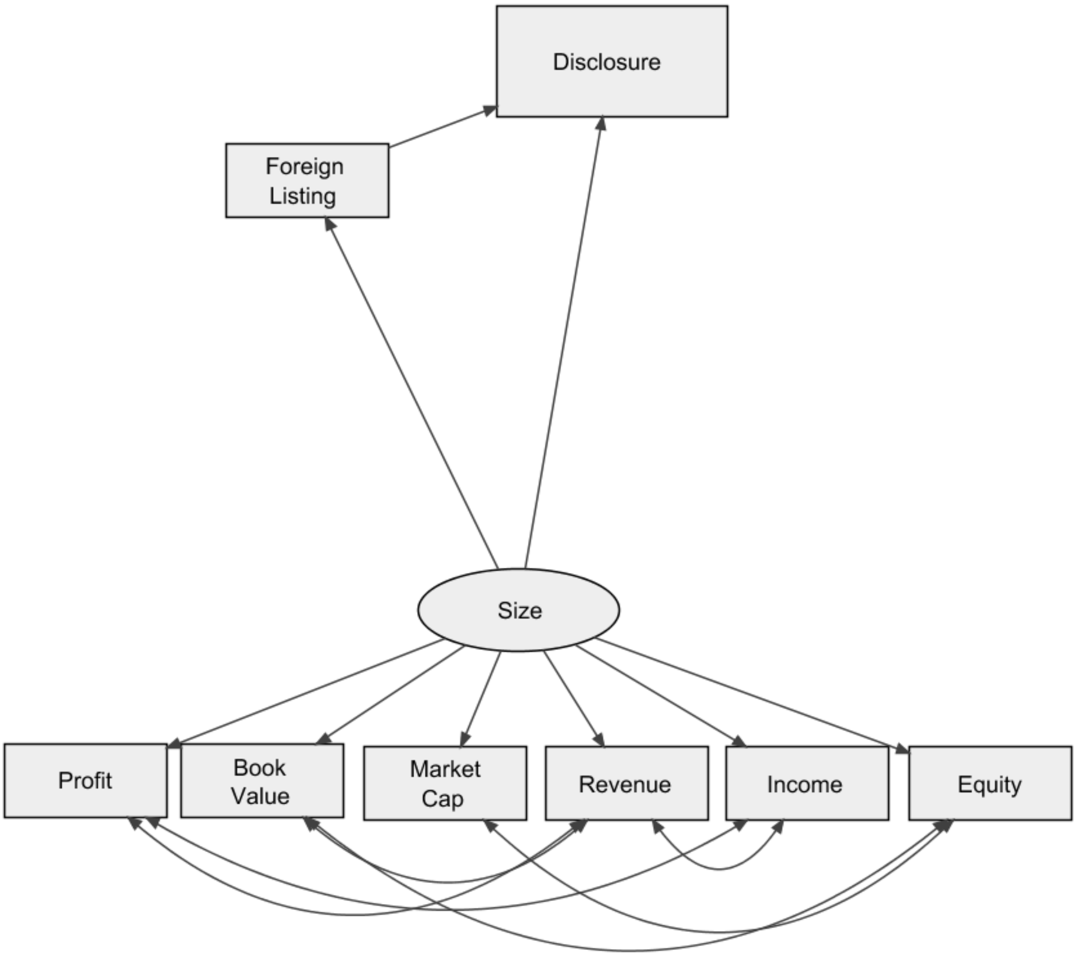

Figure 7.15: Final Model with Normal Score Data

While size has been retained as a cause of disclosure, there is no debt measure. Size and foreign listing may each have agency costs as discussed in chapter 5, but overall this model provides limited information on the theory and is not discussed further. Estimates and fit tables are provided below. Again, no MI-derived changes beyond size covariances were significant.

Table 7.35: Agency Model SEM estimates, normal score data							
			Estimate	S.E.	C.R.	P	Std. Estimate
Foreign	<---	Size	0.374	0.024	15.628	0.000	0.412
Equity	<---	Size	1				0.858
Income	<---	Size	0.729	0.028	25.589	0.000	0.625
Revenue	<---	Size	0.919	0.026	35.901	0.000	0.805
Market Cap	<---	Size	1.028	0.024	43.614	0.000	0.889
Book Value	<---	Size	1.064	0.018	58.201	0.000	0.914
Profit	<---	Size	0.656	0.029	22.449	0.000	0.563
Following	<---	Size	5.929	0.183	32.35	0.000	0.762
Following	<---	Foreign	0.758	0.162	4.681	0.000	0.089

Table 7.36: Agency Model Fit, Normal Score Data								
Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
315.935	13	0	24.303	0.968	0.934	0.45	0.127	0.005

The latent correlations for the size latent in this model as usual show strong correlations with the indicator variables.

Table 7.37: Latent Correlations	
Latent	Size
Foreign listing	.394**
Profit	.574**
Book Value	.958**
Market Cap	.934**
Revenue	.833**
Income	.630**
Equity	.908**
Following	.761**

7.5: Legitimacy

The final set of models is based on Legitimacy Theory. The initial model is:

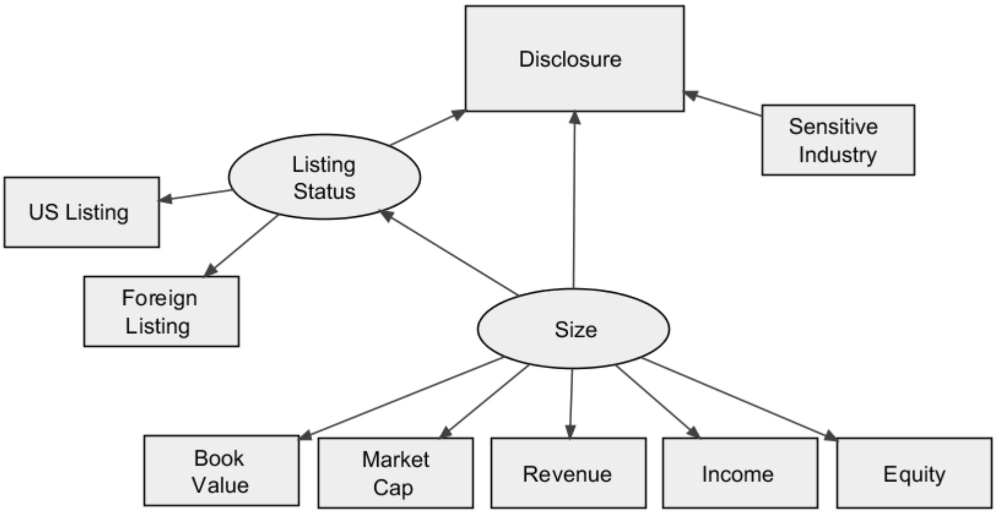

Figure 7.16: Legitimacy Theory, Initial Model

The whole model is based on the idea of negative attention being prevented by disclosures pointing out why it is not warranted. The included variables are all items that could lead to such. The critical one here is the Sensitivity variable. Size is included because of the PCH aspect of the model, as Watts and Zimmerman (1978) argued that larger companies would be subject to greater political pressure. Listing is considered related to size, but may also indicate a globalised company, which is sometimes considered a negative trait.

7.5.1: Base data

When tested with the base data, the final iteration of the model is:

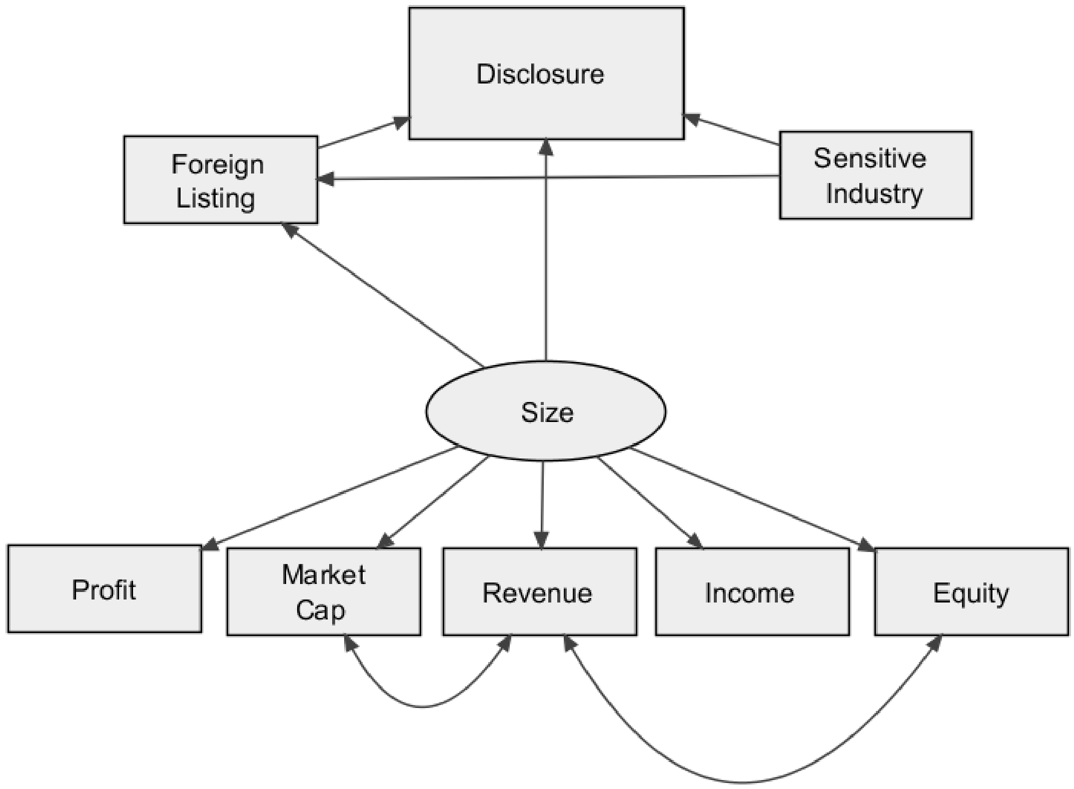

Figure 7.17: Legitimacy Theory, Final Model

The main MI-derived change here is that of allowing sensitivity to cause foreign listing. There is theory derived logic to this connection. Assuming that foreign listing is reflective of foreign operations, the multiple listing decision may allow locals easier access to the company's decision-makers processes and defuse possible tensions between the locals affected by company actions and its distant headquarters.

There are two differences compared to the original model pictured in figure 7.16. First, as usual the listing latent was reduced to a single measure. The second is that it emerged that the Book Value of the firm was not making a meaningful contribution to this model, but adding profit as a measure of size was of some value.

Once again, there is little to be said in terms of reliability and validity as the various tests are either identical to previous models or differ only to such small extents as to be practically identical.

Table 7.38: Legitimacy Theory Estimates

	Estimate	S.E.	C.R.	P	Std. Estimate
Foreign <--- Size	0.000	0.000	3.823	0.000	0.102
Equity <--- Size	3.363	0.159	21.214	0.000	0.973
Income <--- Size	0.930	0.044	21.356	0.000	0.993
Revenue <--- Size	2.514	0.128	19.708	0.000	0.801
Market Cap <--- Size	8.094	0.379	21.366	0.000	0.995
Profit <--- Size	1.000			0.000	0.493
Following <--- Size	0.000	0.000	4.701	0.000	0.117
Following <--- Sensitive	1.072	0.319	3.355	0.000	0.081
Following <--- Foreign	3.187	0.206	15.493	0.000	0.375

The standardised regression weights highlight that the theory is not incorrect, but neither is it powerful. Sensitivity, the core of this theory, is by some margin the weakest explanation of disclosure. This suggests that Legitimacy Theory has some explanatory power, but is not the main explanation of disclosure.

Table 7.39: Legitimacy Theory Fit

Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
785.075	19	0.000	41.320	0.937	0.909	0.636	0.168	0.408

While the basic indices suggest good fit, little else does. The RMSEA is approximately double the maximum recommended value and the model itself is highly significant, indicating poor fit. The Bollen-Stine p-value does, however, indicate a moderate level of fit.

Correlations involving the size latent variable in this model show the expected low correlations with non-indicators and high correlations with indicators. As in the other base data models, one indicator has a low correlation.

Table 7.40: Latent Correlations	
Latent	Size
Foreign listing	.104**
Sensitive	0.14
Profit	.493**
Market Cap	.998**
Revenue	.804**
Income	.995**
Equity	.974**
Following	.158**
Foreign listing	.104**

7.5.2: Outliers Removed

Modelling Legitimacy Theory after removing the outliers from the sample results in the model:

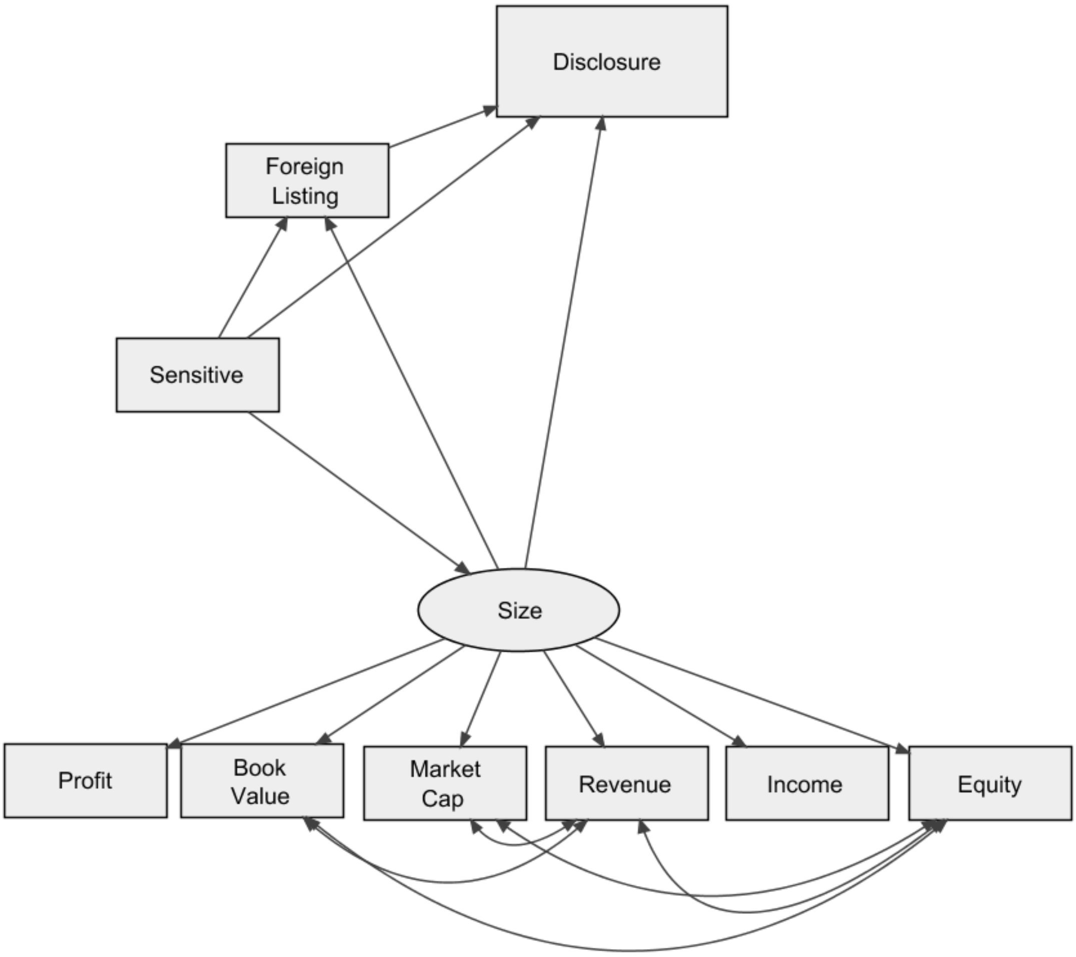

Figure 7.18: Final Model with Outliers Removed

MI examination suggests this model behaves much the same as the full data version, but with one difference. Like the full data Legitimacy model, foreign listing is suggested (and allowed) as a cause of size and covariances among size indicators are allowed. The MIs also suggest allowing sensitivity to cause foreign listing.

Table 7.41: Legitimacy Model SEM estimates, outliers removed							
			Estimate	S.E.	C.R.	P	Std. Estimate
Size	<---	Sensitive	339305	71323.95	4.757	0.000	0.128
Foreign	<---	Size	0.000	0.000	6.948	0.000	0.183
Foreign	<---	Sensitive	0.290	0.040	7.245	0.000	0.189
Equity	<---	Size	1				0.759
Income	<---	Size	0.440	0.010	42.311	0.000	1.004
Revenue	<---	Size	1.677	0.042	40.07	0.000	0.751
Market Cap	<---	Size	3.495	0.076	46.006	0.000	0.924
Book Value	<---	Size	3.404	0.163	20.824	0.000	0.470
Profit	<---	Size	0.356	0.01	36.056	0.000	0.868
Following	<---	Sensitive	0.462	0.306	1.513	0.130	0.036
Following	<---	Size	0.000	0.000	13.42	0.000	0.338
Following	<---	Foreign	2.381	0.202	11.777	0.000	0.288

Sensitivity does influence disclosure, but the effect is very weak. Other variables are much more powerful and this suggests a low power to Signalling Theory.

Table 7.42: Legitimacy Model Fit, outliers removed								
Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
812.946	19	0.000	42.787	0.914	0.840	0.482	0.175	0.010

Unusually, this fit is actually overall worse than that of the base data.

Correlations involving the size latent are presented below. As is normal for the outlier-free data, the correlation with book value is low for an indicator variable while the correlation with analyst following is high for a non-indicator.

Table 7.43: Latent Correlations

Latent	Size
Foreign listing	.211**
Sensitive	.128**
Profit	.874**
Book Value	.471**
Market Cap	.934**
Revenue	.758**
Income	.989**
Equity	.762**
Following	.416**

7.5.3: Logarithmic Data

Using logarithmic data provides a similar model, implying that this model is relatively invariant to data transformations.

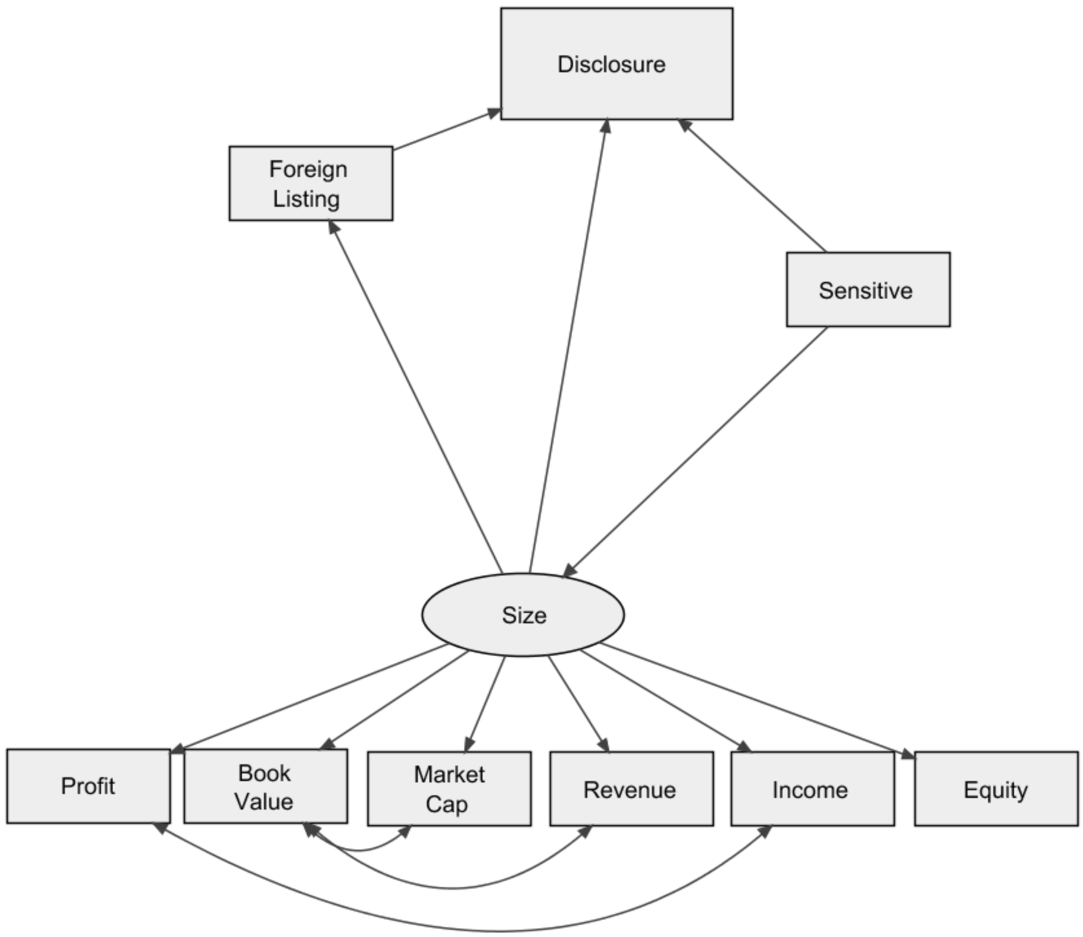

Figure 7.19: Final Model with Logarithmic Data

The MI-derived changes are the usual size covariances and a link from sensitivity to size. This was allowed for the possible fit improvement rather than any clear theoretical reason. A company increasing its scale in response to sensitivity is not a clear legitimation strategy. Instead, it is as argued in other models, that sensitivity is reflective of industry sector, and this defines the market conditions and growth potential for the company

Table 7.44: Legitimacy Model SEM estimates, logarithmic data							
			Estimate	S.E.	C.R.	P	Std. Estimate
Size	<---	Sensitive	0.220	0.040	5.526	0.000	0.150
Foreign	<---	Size	0.439	0.028	15.525	0.000	0.411
Equity	<---	Size	1				0.778
Income	<---	Size	0.247	0.023	10.968	0.000	0.293
Revenue	<---	Size	2.633	0.066	39.708	0.000	0.918
Market Cap	<---	Size	3.212	0.103	31.291	0.000	0.763
Book Value	<---	Size	3.381	0.084	40.293	0.000	0.929
Profit	<---	Size	0.140	0.01	13.403	0.000	0.356
Following	<---	Sensitive	0.440	0.209	2.111	0.035	0.033
Following	<---	Size	7.344	0.224	32.785	0.000	0.804
Following	<---	Foreign	0.555	0.146	3.797	0.000	0.065

As before, the sensitivity effect is weak and other variables explain more.

Table 7.45: Legitimacy Model Fit, logarithmic data								
Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
288.977	22	0	13.135	0.961	0.94	0.587	0.092	0.085

Again, the fit is actually better in the base data model.

Correlations for the size latent are presented below. As with other logarithmic data models, two indicators have low correlations.

Table 7.46: Latent Correlations	
Latent	Size
Foreign listing	.400**
Sensitive	.146**
Profit	.818**
Book Value	.929**
Market Cap	.788**
Revenue	.938**
Income	.408**
Equity	.325**
Following	-.067*

7.5.4: Normal Score Data

Finally, using normal score data once again provides a similar model:

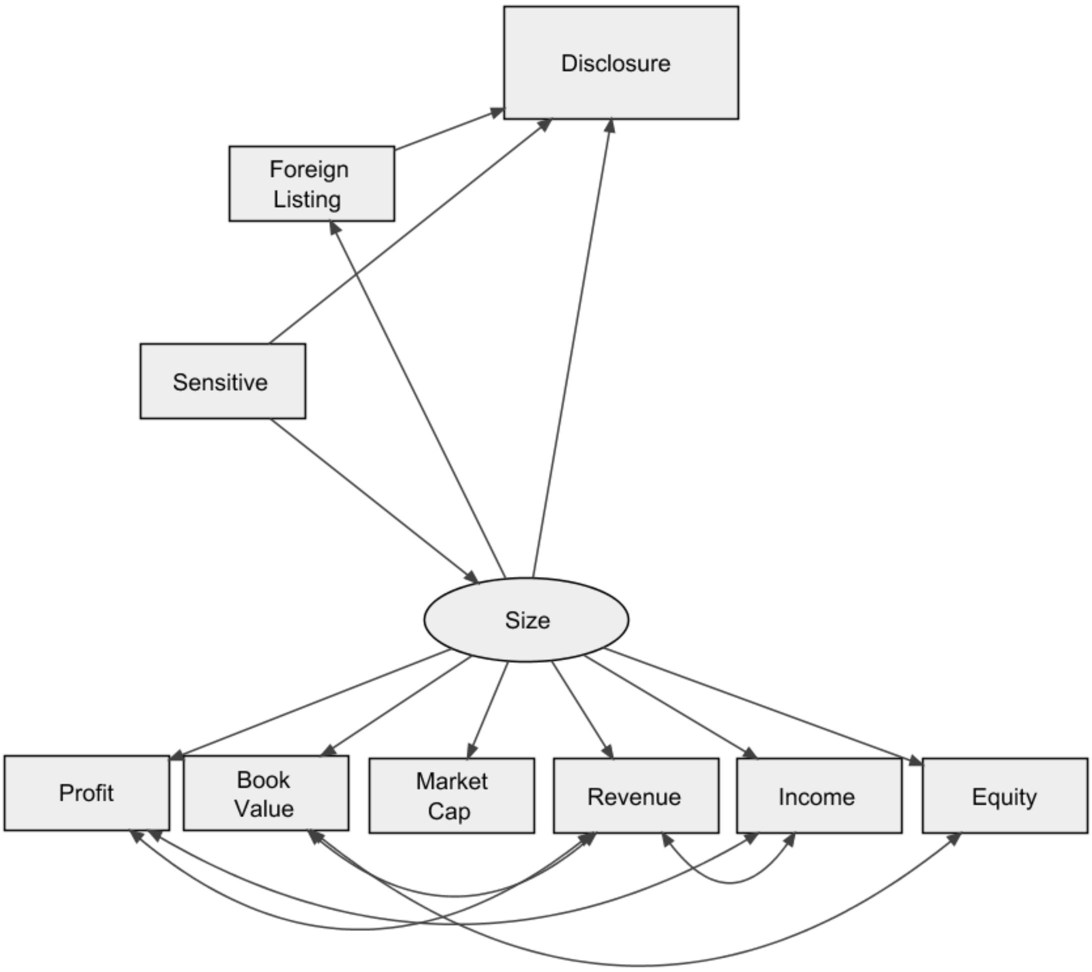

Figure 7.20: Final Model with Normal Score Data

MI-derived additions to this model are mostly identical to those for the logarithmic Legitimacy model and discussion is not repeated.

Table 7.47: Legitimacy Model SEM estimates, Normal Score Data

			Estimate	S.E.	C.R.	P	Std. Estimate
Size	<---	Sensitive	0.205	0.048	4.267	0.000	0.115
Foreign	<---	Size	0.338	0.022	15.159	0.000	0.387
Equity	<---	Size	1				0.891
Income	<---	Size	0.736	0.025	29.379	0.000	0.656
Revenue	<---	Size	0.941	0.02	46.49	0.000	0.856
Market Cap	<---	Size	0.952	0.021	46.404	0.000	0.855
Book Value	<---	Size	1.069	0.018	59.858	0.000	0.953
Profit	<---	Size	0.664	0.026	25.421	0.000	0.592
Following	<---	Sensitive	0.803	0.227	3.533	0.000	0.060
Following	<---	Size	5.439	0.158	34.515	0.000	0.728
Following	<---	Foreign	0.922	0.158	5.85	0.000	0.108

The effect of sensitivity is weak here, but higher than in other models for the theory.

Table 7.48: Legitimacy Model Fit, Normal Score Data

Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
436.835	20	0	21.842	0.957	0.926	0.532	0.121	0.005

As with other models, the fit of this one is worse than that of the base data.

Correlations involving the size latent are presented below. As in other normal score models, the indicators correlate strongly.

Table 7.16: Latent Correlations	
Latent	Size
Foreign listing	.393**
Sensitive	.116**
Profit	.623**
Book Value	.962**
Market Cap	.935**
Revenue	.836**
Income	.673**
Equity	.929**
Following	.470**

7.6 Conclusion

Any conclusion will involve examination of fit. Table 7.50 is a repeat of Table 7.1, a series of SEM fits including the regression models. Fit measures are explained in section 7.1 and 6.7.

Table 7.50: Table of SEM Results

Model	Chi-square	DF	P-value	CMIN/DF	NFI	TLI	PNFI	RMSEA	B-S P-value
Regression Models									
Reg 1	318.392	15	0.000	21.226	0.544	0.374	0.389	0.119	0.139
Reg 2	67.262	15	0.000	4.484	0.741	0.694	0.529	0.049	0.070
Reg 3	166.880	15	0.000	11.125	0.673	0.567	0.482	0.084	0.174
Exploratory Models									
Base	2177.565	24	0.000	90.732	0.841	0.764	0.561	0.250	0.005
Outlier	1180.702	65	0.000	18.165	0.899	0.865	0.642	0.112	0.005
Log	872.200	28	0.000	31.150	0.882	0.816	0.549	0.145	0.035
Normal	3122.814	46	0.000	67.887	0.791	0.703	0.551	0.216	0.050
Agency Theory Models									
Base	4.797	3	0.187	1.599	0.984	0.987	0.492	0.020	0.055
Outlier	918.821	54	0.000	17.015	0.885	0.891	0.637	0.108	0.010
Log	348.056	27	0.000	12.891	0.954	0.929	0.572	0.091	0.050
Normal	315.935	13	0.000	24.303	0.968	0.934	0.450	0.127	0.005

Table 7.34 continued : Table of SEM Results									
Model	**Chi-square**	**DF**	**P-value**	**CMIN/ DF**	**NFI**	**TLI**	**PNFI**	**RMSEA**	**B-S P-value**
Legitimacy Theory Models									
Base	179.117	16	0.000	11.195	0.986	0.977	0.563	0.084	0.557
Outlier	812.946	19	0.000	42.787	0.914	0.840	0.482	0.175	0.010
Log	288.977	22	0.000	13.135	0.961	0.940	0.587	0.092	0.085
Normal	436.835	20	0.000	21.842	0.957	0.926	0.532	0.121	0.005
Signalling Theory Models									
Base	180.239	23	0.000	7.836	0.986	0.980	0.631	0.069	0.562
Outlier	915.613	44	0.000	20.809	0.921	0.886	0.614	0.120	0.005
Log	312.755	28	0.000	11.170	0.958	0.938	0.596	0.084	0.005
Norm	1661.922	41	0.000	40.535	0.894	0.832	0.555	0.166	0.005

There are conclusions to be drawn without considering theory. Fit is generally better overall among those models taking advantage of SEM. The models named Latent, Signalling, and Legitimacy all contain at least one latent variable and have some of the explanatory variables explaining others. By contrast, the three regression models remade in SEM have lower fits by most measures and by their nature do not allow either latent variables or inter-explanatory dependencies. This suggests that there is extra information on disclosure behaviour provided by the use of SEM, although in this case not enough to generate clear good fit where the regressions failed. The exception to this rule is the Agency model, which is the best fit despite it not taking advantage of SEM at all. The final base data Agency model, as seen in Figure 7.12, has neither latent variables nor relationships between explanatory variables.

The final aspect worth considering is what common elements there are among models. The three regression models are not considered here. They are effectively covered under 6.7 above; despite the different methodology, the each model is effectively the same as the regression of the same name. While there are some minor differences in the exact p-values and estimates, any discussion of variables would be identical to that performed above.

Looking at the other models brings in some new points, although few of them can be connected back to the literature. Most models have multiple listing as the single most powerful variable, having the largest standardised estimate. Sensitive industry membership tends to be the second most powerful.

Size is important in almost all of the models and is always a latent variable of five measures, which in turn suggests that there is not a single clearly correct answer to questions of how to measure this concept. In addition, this variable is in two models allowed as a potential cause of foreign listing, in both cases with significant results although the effect is not strong. Buzby (1975) discussed the possible effects of such a link early in the literature, arguing that Singhvi and Desai's (1971) results could be explained if one exists. However, the limitations of regression mean that it has generally been difficult or impossible to confirm such a connection exists outside of Grüning's (2007) work providing some support for the idea.

Many of the findings about the significance and power of individual variables are similar to those of the regressions, although differences do occur. Some are from variables not included in the final SEMs, although as a rule this indicates that the variable in question was insignificant. No SEM contains volatility in the final formula, in keeping with the literature's mix of findings on this matter. The other largely lacking variable is debt finance, which appears in only the Agency model but is insignificant there. Again, however, this has had very mixed results in literature and the lack of significance here is not surprising.

7.6.1: Conclusions from theory modelling

This section discusses possible interpretations in terms of the theory models.

In terms of the numbers alone, the best fit is on the Agency Theory model with the basic data. It fits well by almost all measures and, uniquely, the chi-square test has an acceptable fit for this model. No other model tested here manages this feat.

However, not all measures favour the model. First, the Bollen-Stine p-value is acceptable, but low. Second, the PNFI is surprisingly low. In itself this is not a major sign of poor fit, but the model is very simple and the PNFI is designed to penalise complex models.

The third and largest problem comes from looking beyond the fit measures. The model tests the power of listing status, financial performance, and debt finance as explanations of disclosure. While listing and performance can be justified in terms of Agency Theory, debt finance is the easiest of the three to justify and has the clearest implications for the theory. The fact that it is insignificant in the final analysis indicates that Agency Theory is not explaining the disclosure behaviour of this sample. Looking at table 7.18 for standardised coefficients, all that can be said is that listing status is reasonably powerful as an explanation of disclosure, but almost all models find listing to be important. The Agency model is the best fit by most measures, but it says little about the theory it is intended to analyse.

We turn now to the second-place model for potential explanation of disclosure. However, there are a few candidates for this model. Regression 2 is an immediate possibility as it has the next lowest RMSEA, but its B-S p-value indicates a low chance that the model is correct and the fit index measures are low.

By RMSEA, the next model to consider is that of Signalling Theory using the basic data. At 0.084, the RMSEA is barely unacceptable (although the default 90% confidence interval calculated by AMOS does range below 0.08). By other measures, this model does well. The indices are both above 0.9, a good sign despite other models doing better by these measures.

Analysis so far looks at single models only, taking a single data type into account. In order to identify the best model across all data types, a simple process is used. Table 7.35 below lists the ranks of each model by data type.

Table 7.35: Theory ranks by data used

	Agency	Legitimacy	Signalling
Base	3	2	1
Outlier	1	3	2
Logarithm	3	2	1
Normal	2	1	3
Total	9	8	7

The final row of the table above totals the rankings for each model. The lowest total (i.e. most consistently well-fitting) is the Signalling Theory model. This is expected following the analysis of each set of models as Signalling models tend to fit consistently well where others are often the best fit by some measures but the worst by others. Signalling Theory is hereafter considered the best explanation of disclosure activity of the three tested, with Legitimacy a close second.

This conclusion does rely on equally weighting all models. Any alternatives involving unequal weights could easily change this ranking and the resulting conclusions. Excluding the unaltered data is a clear option as the other three data sets are all intended to clarify the data, but this has a distinct problem. If this set is removed, all three models would have an equal total rank of 6. The base data set provides a useful tiebreaker in this regard.

Depending on how data is manipulated, all three of the theories tested can be demonstrated to be the single best explanation of disclosure.

If Signalling theory is the primary explanation of disclosure, this presents challenges from a regulatory standpoint. One of the key aspects of the theory is the need for a cost of signals. If signals are free then every entity competing for attention has no real reason to not signal. Different quality levels should still show, but one of the clear signs of a good option in the original theory is the existence of a signal – the entity believed that the cost of signalling was low in relation to the potential later gain. If signals are free, this is true in all cases. Mandatory disclosure is not directly a free signal. There will be costs involved in finding and publishing the information. However, if mandatory then the disclosure package becomes an inevitable cost of doing business instead of a choice the company managers are making. All companies will disclose to some level, meaning that it becomes harder for the true good investments to stand out from the crowd.

The situation becomes worse if the underlying explanation is not Signalling, but Lemons Theory. The two are modelled as one theory here due to similarity of conditions with regard to disclosure in section 4.7. While Signalling is assumed to be the better explanation, there is a possibility that Lemons is the real theory underlying this model. The implications of a uniformity of disclosure under this theory are a loss of quality in the market.

A regulator with the goal of increasing the amount of public information should encourage voluntary disclosure instead of increasing mandatory disclosure requirements.. This approach retains the decision-making usefulness of costly signals.

Legitimacy Theory is the second most consistent model across all data types. However, much like the Signalling models potentially representing Lemons Theory, these models may cover the Political Costs Hypothesis, and the implications change if this is the actual explanation.

According to the results of tables 7.10 through 7.17, size and foreign listing are significant determinants of disclosure as in the Signalling models, and sensitive industry membership is a new significant determinant. In addition, size has a very strong effect on foreign listing. More importantly, the significance of sensitivity provides strong support for the underlying theories.

However, while the modelling overall supports Signalling Theory above other possible explanations of disclosure, the main indicator of this theory (performance measures or latent variables) often has a low standardised estimate for the scale of its effects on disclosure. Models of this theory fit better than others, but the variables included as the main evidence of the theory provide weaker evidence. Sensitive industry membership, the main evidence for Legitimacy Theory, is also consistent and has standardised estimates on a similar scale.

If Legitimacy holds as an explanation, further regulation would be useful. The general rule is that companies with any serious concerns about their public image will reveal information. On the surface, making higher levels of disclosure mandatory would increase the overall level of information in the market and reduce overall information asymmetry. There may even be further-ranging benefits; if potentially controversial companies are disclosing to reduce the questions being asked about their actions, disclosure may soothe tensions that could otherwise lead to protest, creating a public order benefit. However, there is little regulatory gain from this argument; the companies already likely to cause such problems are already disclosing, so higher standards of mandatory disclosure would mostly affect less controversial firms.

If the PCH is instead the explanation, the implications change. This theory holds that, from the company's point of view, mandatory disclosure is to be avoided; companies are voluntarily releasing information in the hopes of preventing regulation. To the regulator, this means there is likely to be some information that may help investor decisions that companies are not releasing. The PCH itself therefore suggests that the regulator should always assume more information could be released, but any attempts to enforce it will be unpopular with companies. The actual information contained in mandatory disclosure may ultimately be of low quality as managers unhappy with the regulation still attempt to release little information.

The implications of a Legitimacy model depend heavily on which of the two possible theories behind it is actually holding. Legitimacy suggests regulation may be useful, but limited in the overall gain since the companies with most benefit from it are already the best at disclosure. The PCH, by contrast, suggests regulation is likely to bring more information to light, but would not be at all popular with the companies affected.

The Agency Theory model using the base data is the best-fitting of those examined in the thesis. This should mean that disclosure is best explained from the starting point of companies trying to reduce potential agency costs. The release of information means two things. First, managers know that their actions are likely to be publicised in some manner. It may simply be that the annual report has a note to the accounts detailing that a given area of the company is unusually costly, but at the very least the financial consequences of their actions are published and can be investigated in more detail. The second useful aspect here is that the information is published, allowing it to be viewed by many and increasing the chances of any irregularities being noticed. In addition to highlighting possible agency costs, this would also make it easier to detect potential fraud within the company provided those collating and publishing the information are not pressured to cover it up.

However, this relies on the model being meaningful. This model came to fit the data very well only after being cut down heavily. Only three explanatory variables remained within the model: Debt/Equity ratio, foreign listing, and return on assets. Of these, Debt/Equity is non-significant, having a p-value of close to 1. This variable is vital to Agency Theory; its insignificance implies that the main cause of disclosure is not related to agency costs. This is not an issue in the outlier-removed and logarithmic Agency models, both of which contain some form of debt finance variable as a significant determinant

There is an Agency explanation in the listing variable. A company with operations outside of its home territory, which is a common reason to list in another location, has distant offices that HQ cannot easily oversee. Agency problems are easy to miss from this alone, plus there may be language barriers that enable easier concealment of managerial inefficiency. However, as addressed in section 5.2.2, there is good reason to expect multiple listing to lead to improved disclosure without considering any sort of monitoring of agency costs. Any agency reduction effects from this variable are at best minor support for the idea of Agency Theory explaining disclosure.

Ultimately, the base data model says only that performance and listing lead to disclosure. This is useful information as dealt with in the previous section, but says little about the theory it was designed to model. The insignificance of debt finance as an explanation of disclosure implies that agency cost reduction is largely not connected to the decision to disclose information.

If one theory is to be selected as a single explanation of disclosure, it is Signalling. The other theories fit better in some cases but Signalling models are consistently among the best fitting regardless of the form of data used. This means the primary conclusions to be drawn are explained above.

However, all three models contain some validity, and the earlier analysis of theory in chapter 4 demonstrates that the three may all coexist and explain disclosure to some extent. There is potentially some explanatory power in all three theories, meaning that the all three of the discussions presented earlier in this section may have implications for disclosure.

7.6.2: Alternative interpretations

While the models are intended to investigate the specific theories discussed above in section 7.6.1, they contain information beyond this. Where it appears in the model, size is almost always a significant and positive determinant of the disclosure measure. Similarly, listing status has near-universal positive effects. These two are consistently important. Membership of a sensitive industry and performance are both frequently significant, although the power of performance tends to be lower. Debt finance and volatility can generally be ignored as insignificant in some manner. Debt finance tends to show a statistically significant positive effect on disclosure, but standardised estimates demonstrate a near-zero level of influence over the dependent variable. Volatility, by contrast, is never statistically significant.

In purely descriptive terms, this suggests that the companies making the best disclosures are large multinationals in sensitive industries. The profitability measure is not so clearly important, though does play a part, while the other two are at best arguably important. The two most important variables both have solid arguments behind them. Size may encourage disclosure because the company's scale of operations means the central HQ requires collated information, and the company can then disclose that information if it sees any advantage in doing so.

Multiple listing leads to the firm having to follow multiple sets of listing rules and may cause a form of unintentional disclosure. The company is likely to make one report covering all required information and therefore publish a little bit of extra information in each jurisdiction. The high power of this variable provides evidence against Biddle and Saudagaran's (1989, 1995) ideas. The paper suggests that companies will select their second and subsequent listing locations in order to minimise the added reporting requirements. For the multiple listing measures to be so powerful indicates that this is not a concern among companies. It is possible that attitudes have changed since the two papers, the latter of which is nearing two decades old. Similarly, the sample may make a difference. This research has used only a UK sample, meaning most companies can potentially gain access to the large US markets with minimal language barriers. It may be the case that, for the sample used, the ease and gains of US listing outweigh the apparent costs of the stricter disclosure requirements. There is the additional possibility that the relationship is reversed; a company which is willing to disclose more information is more willing to list in multiple locations, not considering the extra disclosure requirements to be a problem.

The positive effect from performance results from manager interests; low results lead to less detailed information in order to prevent any specific manager(s) being blamed, while high performance means the managers can show off their talent with extra information.

Sensitivity is harder to explain without reference to Legitimacy Theory or the Political Costs Hypothesis. Sensitivity leading to disclosure suggests that potentially volatile public opinion can be settled with the release of information, which is exactly what Legitimacy theory predicts. However, reputation concerns beyond legitimate behaviour may provide an explanation. All companies need to maintain a good reputation in order to keep a steady flow of customers and the income they bring. If the firm deals with the general public, giving information out publicly may be a means to this end. Taking the argument further, there may be a reputation risk to the company and this has cost of capital implications. From an investor's point of view, any damage to the firm's reputation will harm its ability to generate returns, adding a risk to the investment. This is factored in to the investor's calculations by requiring a reputation risk premium on any investment made in the company, similar to Lang and Lundholm's (1996) information risk premium. Disclosure to the general public may indirectly reduce this by stabilising the company's reputation in the overall market.

Chapter 8: Conclusion

There are four sections to this chapter. Section 8.1 discusses and summarises the answers to the research questions set out in chapter 1. Section 8.2 covers the meaning of the outcomes for disclosure. Section 8.3 discusses the identified limitations of the research, and section 8.4 discusses possible future research based on the ideas contained in this project.

8.1: Summary of Findings

Three main aspects of contribution were explained in chapter 1. There were the triangulation of results using another method, the ability to model theory more directly than regression, and a possible improvement in model fit from the inclusion of causal links among explanatory variables. The theory modelling aspect is explained in section 8.2 below, while triangulation is covered in this section.

The results of the models presented here largely help to confirm previous findings. Most models find a large effect on disclosure from the company's listing status and size, as is consistent in other papers. This supports the idea that Singhvi and Desai (1971) and following papers have correctly identified the two major causes of disclosure. Further triangulation of results using additional samples of different methods would still be useful.

The use of SEM gave the possibility of allowing explanatory variables to be linked together, often with causal links between some where there is reason to expect such effects. As discussed in chapter 1, the intention behind this was both to better model theory and to improve model fit by including effects not previously considered. If the fit of the models was an improvement over regression models then the low fits observed in literature may be explained by the models not capturing all effects of the variables included.

In this regard, results were mixed. SEM fits are not directly comparable to regression fits due to the lack of any single measure of fit in SEM, as discussed in chapter 5. However, regression models have been rebuilt as structural models in chapter 6, giving some measures of fit for structural models that assume explanatory variable independence. The models in chapter 7 do allow such connections between explanatory variables and offer improvements in fit over the models with independence among explanatory variables. This therefore suggests that the relaxation of the variable independence assumption does improve model fit, but it is difficult to translate the range of SEM fit numbers into an R^2 figure for direct comparison of the improvement.

While this appears to suggest that variable independence is causing problems for regression models of disclosure, the fits of the structural models were rarely in the range where they would be considered clearly well-fitting. This suggests that the models are still missing some important determinants of disclosure. Allowing explanatory variables to be correlated has helped with the fit and therefore improved upon the usual models of disclosure, but it is not the only cause of poor fit in literature to date.

In addition to the contributions to literature, three research questions were set out in the introduction and are directly answered here.

RQ1: What combination of company characteristics drives corporate disclosure?
The models tested in chapters 6 and 7 are in broad agreement on this question. The results of testing each variable are described here.

The most powerful determinant of disclosure in the models tested here is multiple listing. This variable is measured as the total number of countries in which a company is listed in most models, although one model, Regression 2 (detailed in section 6.2.2) uses an alternative measurement that uses US listing alone as a dummy variable. While the structural models could in principle use both measures, the use of both created problems likely due to the inherent overlap; total foreign listing includes US listing.

In all cases, a measure of foreign listing is included in the model and found to have a significant and positive effect on disclosure. Further, comparing the standardised estimates of the significant variables, the estimate for foreign listing is invariably the largest. In the regressions its highest value is 0.350 (regression 3, table 6.3), while the highest value any other variable reaches is 0.257 (revenue, a measure of size, in regression 2, table 6.2). In the structural models, the lowest estimate for total foreign listing is 0.368 (Signalling model, Table 7.10), while the highest estimate for any other variable in the SEMs is 0.159 (for performance in the same model). In the one model that uses US listing as a measurement instead (regression 2, table 6.2), the standardised estimate is much lower at 0.150. The total foreign listings of a company are an important determinant of disclosure practices.

Almost all models find the size of a company, by whichever measure or combination of measures is used, to have a significant and positive effect on disclosure. Although not as powerful as foreign listing, this variable has a long history of positive effects in existing literature. The reasons why size should influence disclosure in this way are well-established, which may create a cyclical effect. The sound reasoning and history of results encourage researchers to use the variable, and common use means new possible explanations are theorised, tested, and accepted.

The sole exception found is the model for Agency Theory with the base data, pictured in Figure 7.12, which lacks any size variable. The initial version of this model (figure 7.11) did not directly connect size to disclosure, arguing that size effects under the theory would be better represented as influences on other variables rather than direct effects. However, even if the connection is made, little changes – if included, the size effect on disclosure has a p-value of 0.516, indicating a highly insignificant variable. Had it been included, the model refinement process would have deleted the connection immediately, resulting in the model that was analysed in chapter 7. This makes the model very unusual; with size being significant in all other models, and a majority of the existing literature, the lack of effect found here is a strong contrast to others.

Financial performance has quite consistent results here despite varied findings in literature. Where included it consistently has a significant and positive effect on disclosure unless logarithmic data is used, in which cases the effect is negative. The size of the effect is less consistent, being dependent on exactly how the variable is measured, as observable from the varied standardised estimates. Table 6.1 uses return on assets and estimates 0.162, table 6.2 uses earnings margin for an insignificant and small estimate, while 6.3 uses profit/size and estimates 0.066. In the structural models, similar estimates are obtained depending on what is included in the model. The exploratory model's final iteration uses only profit/size, obtaining an estimate of 0.067, similar to that of table 6.3. The Signalling model detailed in table 7.10 uses a combination of profit/size and return on assets and obtains a standardised estimate of 0.159 for the effect of performance on disclosure, which is similar to the return on assets estimate in table 6.1 and suggests that this variable is dominating the performance latent.

Debt finance, like financial performance, has a significant and positive effect in most cases where it is included. This variable is an important part of Agency Theory explanations of disclosure in the thesis, but was insignificant in two of the four Agency models.

Sensitive industry membership is found to have a consistently significant effect on disclosure. Where included, the variable's effect is small, with a standardised estimate below 0.1 in almost all cases.

Volatility never has a significant effect. It was not included in some of the theory models as no direct effect explained by the theory was identified. Where it is included in a model, it is never significant at the 5% level. Using the less rigorous 10% level would offer it significance in some cases, e.g. p=0.095 in regression 2. However, examination of the standardised estimates in each table demonstrates that the volatility effect is miniscule.

In answer to the research question, then, multiple listing status is the most important determinant of disclosure, followed closely by the size of the company. Sensitive industry membership has a weaker but consistent effect. Other variables are less powerful; financial performance generally indicates better disclosure, although this is highly variable based on how the idea is measured, while debt and volatility are generally insignificant.

RQ2: Which theory or combination of theories best explains disclosure?

Section 4.7 demonstrates that all of the theories examined here are able to coexist, meaning that each can hold explanatory power over the question of what drives disclosure. A combination of theories is therefore a possible outcome. Lemons and Signalling Theory, after analysis, were considered identical in terms of their effects on disclosure, while Legitimacy and the Political Costs Hypothesis were identical in terms of the models. Differences are identifiable in the two, but the models make it impossible to distinguish between the two. The comparison then leaves three distinguishable theories: Signalling, Agency, and Legitimacy.

The two methods used result in some disagreement over which theories offer the best explanation of disclosure. The regressions support all three of Signalling, Agency, and Legitimacy to some extent. The consistent significance of sensitive industry membership implies strong support for Legitimacy above the others. By contrast, this theory is the second most consistently supported when directly modelled with a SEM approach.

The overall best fitting models across all data types are those for Signalling theory, which include several of the best-fitting structural equation models tested. It is the best supported of the three theories under SEM.

RQ3: Does the use of SEM offer any benefit over regression analysis for the research questions above?

When regression models are remade in SEM form, the results are almost identical and most discrepancies are very small, often at the third decimal place. While direct comparison is not possible, the two model types obtain near-identical results in the same circumstances.

As described above, regression and SEM have different results in terms of the preferred theories. However, the regression models are used to reach a conclusion by examining the implications of which variables are significant and which are powerful. By contrast, the use of SEM enabled the direct modelling of each theory. This method is preferred as it better captures the nature of each theory. SEM, then, has benefits over regression that justify its greater complexity.

However, fit values remain low. As discussed in section 3.6, R^2 values for regressions in the wider literature are often low and suggest that further variables are required for a full explanation of disclosure behaviour. The use of SEM does not provide a comparable fit measure, making direct comparison difficult, but the common finding in the models used here is that the fit is lacking in some way. Allowing correlations among explanatory variables has changed the interpretation of some common variables, but has not been demonstrated to improve the model fit to acceptable levels.

When directly comparing the methods by rebuilding regression models in SEM form, an important difference was noted. Although the two techniques demonstrate similar estimates, transformation of data reveals an important difference. Fit values for regressions are sensitive to the data type used, with the logarithmic and normal score data having large improvements to R^2 values that clearly demonstrate an improved fit when using these alternatives. This was not the case for SEM, which tends to have relatively invariant fit values across data types. At least for the data used here, regression techniques show that a transformation of the data may help to explain variance in disclosure, giving it an advantage over SEM in this regard.

8.2: The outcomes in context

This section asks what the results above mean for disclosure and companies.

Most models have consistently shown that the most disclosure is coming from large, multiple listed firms with strong recent performance. From a regulatory view, any attempts to mandate further disclosures must therefore be targeted towards the opposite – smaller, localised companies, often with weaker performance. Some of the reasons these variables are included indicate that any targeting of these companies would be problematic, however. Larger companies are more likely to have the resources necessary to disclose at all, plus an economy of scale is likely to exist and favour larger firms in disclosure. Multiple listing tends to mean foreign operations and therefore a need to gather and collate information for internal control reasons, which can be disclosed at a lower marginal cost. High profitability ratios suggest the firm has some extra resources available in the immediate term. In each case, firms that have lower measures of each have either less need to ready the information internally or be less able to afford any costs involved in gathering it. Such targeting is unlikely to be popular among companies as it places relatively greater burden on smaller businesses.

8.3: Limitations

There are a number of limitations in the research performed.

Disclosure as a concept is difficult to measure. There is the question of whether to measure quality or quantity of disclosure which complicates the matter further, but even using only quantity measurements raises challenges. One of the common approaches, as discussed in chapter 5, is to create an index of disclosures and score companies based on how many of them are included. However, a variation on this approach allows for finer measurement, such as allowing one point if an item is included and another if it is discussed thoroughly. Further, there is some researcher judgement required as to whether an item is included sufficiently in order to qualify for a point on the index or not.

The use of a proxy measure in any context will cause problems. By nature, a proxy is related to the item of interest but is not a direct measurement of the item. It is therefore possible for a proxy to vary for reasons unrelated to the underlying item. The use of analyst following as a proxy for disclosure is not an exception. It is correlated with disclosure quantity, an important quality in a proxy. Further, it is often used in the index-based approaches described in section 5.2.7; the first step in forming an index is to decide what to include, which often involves surveys of what information is useful to people using the company's information, a survey usually directed towards those using such information in a professional capacity to examine the company, i.e. analysts. Even where external disclosure ratings for companies are used, such as the AIMR rankings, these are commonly formed with analyst consultation.

However, Lang and Lundholm (1993) argue that analysts are driven by market forces. This is central to the argument for the proxy; analysts tend to follow those companies that release enough information to be easy to follow, minimising the analysts' own information-gathering time. However, the same market forces may cause analysts to follow companies for other reasons. Once the analyst has finished analysis of a company, the resulting information can be sold to many customers without depleting any stock the analyst has, by the nature of the economics of information. This means there is potentially great profit to the analyst in following a company popular with the analyst's customers even if its disclosures are minimal and require much of the analyst's time to process. Similarly, a company with little information demand is unlikely to be profitable to follow even if its disclosures are thorough. In addition, there is no guarantee that detailed disclosures that would score highly on an index measure translate to ease of analysis; by making more information available, the company may become difficult to analyse through the sheer quantity of information or even obscure relevant information among unnecessary details.

In terms of the results obtained, regression and SEM support different theories. Regression suggests a Legitimacy Theory explanation for disclosure above the others examined. The Legitimacy model is the worst fitting of the structural models, which suggest Signalling Theory is the most accurate. While there is some disagreement in this matter despite the two methods demonstrating compatibility in chapter 6, it is not problematic. SEM has provided the ability to model theories directly rather than identifying which theories best represent the results of a less flexible regression model. However, there may be alternative modelling approaches available that would provide more useful modelling ability. SEM was identified from a few appearances in accounting and finance journals and is not a method in widespread use in the discipline. Most of the information on SEM used in the project has come from marketing and psychology journals, disciplines that have been using the method for decades. Alternative approaches may be available under similar circumstances, common in another discipline not considered here but largely unknown in accounting and finance.

The theories examined here represent a limited range of the possible explanations for voluntary disclosure. Each was selected because it has appeared repeatedly in literature as a possible explanation, but also because each is a long-standing theory in accounting or finance. Each one has broad implications beyond any phenomena or behaviour for which they were originally formed. However, other explanations have been advanced to explain disclosure, often forming new theories for the specific purpose of explaining disclosure. In some cases, the implications of these disclosure-specific theories may be contained within the broad ones studied. As an example, one idea holds that disclosure increases the information available about the company and reduces the risks of something completely unknown to investors causing harm, which in turn reduces an information risk premium investors demand in their expected returns. By disclosing, then, the company's cost of capital is reduced. This theoretical idea may be explained through the broader Signalling Theory, however; the company obtains a benefit in the market by showing more information to investors. While the theories studied will cover many of these more specific explanations, there may be some that do not neatly fit with a broad theory. Further, narrowing the focus onto these more specific explanations may change the models and potentially reveal further insights into when disclosure occurs.

In addition, chapter 4 introduced five theories, but chapter 7 modelled only three. Lemons and Signalling were considered identical in terms of disclosure and there is little that can be done to change this; they make the same predictions due to conceptual overlap. By contrast, Legitimacy and the PCH were given a common model as much because of similar conditions as the practical matter of the models being identical. The problem lies in the selection of variables in the project; alternatives, or the addition of other variables, may make a difference detectable.

The form of data used may influence the results. The single best fit across all models is the Agency Theory model using the unaltered data set, but as discussed in section 7.4.1, this model uses few variables and provides little information on any theory. The next model considered is the Signalling Theory model using the unaltered data, and the third is Legitimacy Theory again using the same data.

If the unaltered data is removed from consideration, however, the conclusion is less clear. The three best-fitting models are all removed with the data, making the remaining models weaker on average. Further, in Table 7.35, the removal of the first row of data would lead to all three theories having a total rank of 6, making it impossible to identify a most convincing explanation of disclosure through this method.

Looking at the fit values in tables 7.1 or 7.34 and excluding rows labelled “base” offers a different means to identify a best-fitting model while excluding the unaltered data. The lowest RMSEA is 0.084 on the logarithmic Signalling model, putting this near to the required maximum of 0.08. Further, this has the lowest χ^2/DF ratio, further suggesting it is the best fit (with the ratio value of 11.170, this is indicative of best fit relative to other tested models and not good fit in absolute terms). The closest competitors is the logarithmic Agency model, which has slightly worse values for most measures but does fit well according to the B-S p-value. Beyond this, the next best fit comes from the logarithmic Legitimacy model, which has marginally better fit index values than the logarithmic Signalling model but is worse on other values. This indicates that the logarithmic data provides the best fit outside of the unaltered data, and both of these data sets rank the theory models the same way.

The conclusion that Signalling Theory best explains disclosure (at least in this sample) is based on Signalling models being consistently good, if not always the best, fits among the models tested. With a convincing reason to use a specific form of data for this testing, this conclusion could easily be changed.

The final limitation considered here is a broad methodological one. To summarise, is this question best studied by large sample statistical methods? While a quantitative approach produces generalizable results, there is merit in a qualitative approach of asking financial managers or members of the board of directors about their disclosure practices. The research performed supports the idea that disclosure is intended to signal the company's best points and stand out against competition for finance, but cannot explain what series of decisions were involved. Another question remains unanswered: What effect are managers intending to achieve with their disclosure practices?

Despite the number of quantitative papers in voluntary disclosure research, qualitative methods have value in the area. As an example, Armitage and Marston's (2008) series of interviews "makes it clearer what [financial directors] are saying about the purpose of disclosure" (p332), providing a useful alternative means of examining why voluntary disclosure occurs. Although not employed in this thesis, the value of qualitative research for determining the reasons for voluntary disclosure is recognised.

8.4: Further research opportunities

Two research opportunities come from considering alternatives. A UK-based sample was used due to researcher familiarity with the sample. However, this limits the conclusions drawn to the UK. Even assuming this does generalise to countries with similar accounting systems, the Anglo-American accounting system is based on shareholders as a source of finance instead of institutional investors in the system used throughout Continental Europe. The first research opportunity is to use the same method with a sample drawn from another country to determine which theories of disclosure best explain observed disclosure practices elsewhere in the world. Some literature (e.g. Zarzeski 1996, Adams et al 1998) uses samples drawn from multiple countries to compare different national practices in one paper. A similar approach would be possible, but a limit of SEM becomes relevant. SEM cannot easily deal with large numbers of dummy variables in the same way as regression. A comparative study with many countries, such as Archambault and Archambault's (2003) comparison across 33 nations, would be more challenging to perform using SEM compared to regression.

The second opportunity is to limit the sample rather than expand on it. For this project, a large sample was selected to ensure statistical power when using an unusual methodology, but SEM does not require a sample as large as that used here. During the course of the project, a pilot study was performed on a sample drawn from the FTSE 350. This pilot study had further data requirements compared to the final study and availability limited the sample to 289 companies without problem. The results of the pilot study are not detailed here, and they are not comparable to the full results due to the pilot being used to refine the approach and different data being used. It is mentioned only to highlight a directly comparable example in which the sample was limited in some way being used with the same methods. Patten's (1992) paper is a good example of this idea as not only is it restricted to the oil industry, it mostly focuses on a few companies in the industry. While the pilot study example would show what occurs among only larger UK quoted companies, other possible limits are readily available.

Additional opportunities for further study emerge from some of the limitations above. One of the limits is that only five theories of disclosure were selected for study. As discussed above, many other explanations have been advanced over time. While this project involved only broad theories that explain various observations in accounting and finance that include disclosure, the disclosure-specific explanations also need analysis. While this has been performed in other literature, the SEM approach used here enables modelling the implications of each explanation directly.

From the third limitation, there are other explanations of disclosure available. It may be worth examining these in the same manner, testing them in comparison to theories. The theory comparison phase would be a potential challenge as these are not generally formalised theories in the same way as those tested here and therefore are less clearly defined (and definable) in terms of necessary and sufficient conditions.

Further, the use of one model for both Legitimacy Theory and the PCH was, as mentioned above and in Chapter 4, due to the differences between the models being indistinguishable using the set of variables involved in the project. This gives impetus to the idea of using a different set of measurements or adding additional variables. Even without this issue, the use of additional variables may offer further information on what drives disclosure. The main challenge involved is identifying suitable variables, as results in literature tend to be mixed for many possibilities.

The final limitation listed in section 8.3, the question of what a qualitative approach may bring to the research question of what drives disclosure, suggests a completely different form of research into the same area. Obtaining information on disclosure intentions is likely to involve qualitative methods, such as Armitage and Marston's (2008) interviews with financial managers, although a large-scale survey of those making the decisions of what is disclosed may be possible. However viewpoints are gathered, it may be useful to compare the results of this hypothetical research to those of the common quantitative literature to determine whether stated intentions match what modelling methods are showing and investigate any discrepancies. An understanding of how disclosure decisions are made may provide further variables to include in models informed by the process of gathering views.

References

Aaker, D.A., Bagozzi, R.P., 1979. *Unobservable variables in structural equation models with an application in industrial selling*. Journal of Marketing Research XVI, 147–158.

Aboody, D., Kasznik, R., 2000. *CEO stock option awards and the timing of corporate voluntary disclosures.* Journal of Accounting and Economics 29, 73–100.

Abraham, S., Cox, P., 2007. *Analysing the determinants of narrative risk information in UK FTSE 100 annual reports.* The British Accounting Review 39, 227–248.

Abrahamson, E., Amir, E., 1996. *The information content of the president's letter to shareholders.* Journal of Business Finance & Accounting 23, 1157–1182.

Adams, C.A., Hill, W.-Y., Roberts, C.B., 1998. *Corporate social reporting practices in Western Europe: legitimating corporate behaviour?* The British Accounting Review 30, 1–21.

Ahmed, K., Courtis, J.K., 1999. *Associations between corporate characteristics and disclosure levels in annual reports: a meta-analysis*. The British Accounting Review 31, 35–61.

Akerlof, G.A., 1970. *The market for "lemons": quality uncertainty and the market mechanism.* The quarterly journal of economics 84, 488–500.

Alford, A., Jones, J., Leftwich, R., Zmijewski, M., 1993. *The relative informativeness of accounting disclosures in different countries.* Journal of Accounting Research 183–223.

Ali, A., Chen, T.-Y., Radhakrishnan, S., 2007. *Corporate disclosures by family firms.* Journal of Accounting and Economics 44, 238–286.

Al-Tuwaijri, S.A., Christensen, T.E., Hughes Ii, K.E., 2004. *The relations among environmental disclosure, environmental performance, and economic performance: a simultaneous equations approach.* Accounting, organizations and society 29, 447–471.

Anderson, J.C., Gerbing, D.W., 1988. *Structural equation modeling in practice: A review and recommended two-step approach.* Psychological bulletin 103, 411.

Anderson, J.C., Gerbing, D.W., 1982. *Some methods for respecifying measurement models to obtain unidimensional construct measurement.* Journal of Marketing Research XIX, 453–460.

Anderson, J.C., Narus, J.A., 1990. *A model of distributor firm and manufacturer firm working partnerships.* The Journal of Marketing 54, 42–58.

Archambault, J.J., Archambault, M.E., 2003. *A multinational test of determinants of corporate disclosure.* The International Journal of Accounting 38, 173–194.

Armitage, S., Marston, C., 2008. *Corporate disclosure, cost of capital and reputation: Evidence from finance directors.* The British Accounting Review 40, 314–336.

Arya, A., Glover, J., Mittendorf, B., Narayanamoorthy, G., 2005. *Unintended consequences of regulating disclosures: The case of Regulation Fair Disclosure.* Journal of Accounting and Public Policy 24, 243–252.

ASB, 2006. *Reporting Statement: Operating and Financial Review.*

ASB, 2007a. *A Review of Narrative Reporting by UK Listed Companies in 2006.*

ASB, 2007b. *A Review of Narrative Risk Reporting by UK Listed Companies in 2006.*

ASB, 2009. *Rising to the Challenge: A Review of Narrative Reporting by UK Listed Companies.*

ASB, 2014. *Guidance on the Strategic Report.*

Ascioglu, A., Hedge, S.P., McDermott, J.B., 2005. *Auditor compensation, disclosure quality, and market liquidity: Evidence from the stock market.* Journal of Accounting and Public Policy 24, 325–354.

Baginski, S.P., Hassell, J.M., Kimbrough, M.D., 2002. *The effect of legal environment on voluntary disclosure: Evidence from management earnings forecasts issued in US and Canadian markets.* The Accounting Review 77, 25–50.

Bagozzi, R.P., 1981. *Evaluating structural equation models with unobservable variables and measurement error: a comment*. Journal of Marketing Research XVIII, 375–381.

Bagozzi, R.P., Yi, Y., 2012. *Specification, evaluation, and interpretation of structural equation models.* Journal of the Academy of Marketing Science 40, 8–34.

Baik, B., Billings, B.K., Morton, R.M., 2008. *Reliability and transparency of non-GAAP disclosures by real estate investment trusts (REITs).* The Accounting Review 83, 271–301.

Bailey, W., Karolyi, G.A., Salva, C., 2006. *The economic consequences of increased disclosure: Evidence from international cross-listings.* Journal of Financial Economics 81, 175–213.

Ball, R., Foster, G., 1982. *Corporate financial reporting: A methodological review of empirical research.* Journal of accounting Research 20, 161–234.

Barker, R., Imam, S., 2008. *Analysts' perceptions of "earnings quality."* Accounting and Business Research 38, 313–329.

Barrett, M.E., 1975. *Annual Report Disclosure: Are American Reports Superior?* Journal of International Business Studies 6, 15–24.

Barrett, P., 2007. *Structural equation modelling: Adjudging model fit*. Personality and Individual differences 42, 815–824.

Barron, O.E., Kile, C.O., O'Keefe, T.B., 1999. *MD&A quality as measured by the SEC and analysts' earning forecasts*. Contemporary Accounting Research 16, 75.

Barth, M.E., Schipper, K., 2008. *Financial reporting transparency*. Journal of Accounting, Auditing & Finance 23, 173–190.

Beattie, V., 2005. *Moving the financial accounting research front forward: the UK contribution*. The British Accounting Review 37, 85–114.

Beattie, V.A., Pratt, K., Scotland, I. of C.A. of, 2002. *Voluntary annual report disclosures: what users want*. Institute of Chartered Accountants of Scotland Edinburgh.

Beattie, V., McInnes, B., Fearnley, S., 2004. *A methodology for analysing and evaluating narratives in annual reports: a comprehensive descriptive profile and metrics for disclosure quality attributes*, in: Accounting Forum. Elsevier, pp. 205–236.

Beattie, V., Thomson, S.J., 2007. *Lifting the lid on the use of content analysis to investigate intellectual capital disclosures*, in: Accounting Forum. Elsevier, pp. 129–163.

Beaver, W.H., 1968. *The information content of annual earnings announcements*. Journal of accounting research 6, 67–92.

Benston, G.J., 1969. *The value of the SEC's accounting disclosure requirements*. Accounting Review 44, 515–532.

Bentler, P.M., 2007. *On tests and indices for evaluating structural models*. Personality and Individual Differences 42, 825–829.

Bentler, P.M., Bonett, D.G., 1980. *Significance Tests and Goodness of Fit in the Analysis of Covariance Structures*. Psychological Bulletin 88, 588–606.

Beretta, S., Bozzolan, S., 2004. *A framework for the analysis of firm risk communication.* The International Journal of Accounting 39, 265–288.

Beretta, S., Bozzolan, S., 2008. *Quality versus quantity: the case of forward-looking disclosure.* Journal of Accounting, Auditing & Finance 23, 333–376.

Bergman, N.K., Roychowdhury, S., 2008. *Investor sentiment and corporate disclosure.* Journal of Accounting Research 46, 1057–1083.

Beyer, A., Cohen, D.A., Lys, T.Z., Walther, B.R., 2010. *The financial reporting environment: Review of the recent literature.* Journal of accounting and economics 50, 296–343.

Beyer, A., Guttman, I., 2012. *Voluntary disclosure, manipulation, and real effects*. Journal of Accounting Research 50, 1141–1177.

Bhamornsiri, S., Schroeder, R.G., 2004. *The disclosure of information on derivatives under SFAS No. 133: Evidence from the Dow 30.* Managerial Auditing Journal 19, 669–680.

Bhushan, R., 1989. *Firm characteristics and analyst following*. Journal of Accounting and Economics 11, 255–274.

Biddle, G.C., Saudagaran, S.M., 1989. *The Effects of Financial Disclosure Levels on Firms' Choices among Alternative Foreign Stock Exchange Listings*. Journal of International Financial Management & Accounting 1, 55–87.

Bloomfield, R.J., Wilks, T.J., 2000. *Disclosure effects in the laboratory: Liquidity, depth, and the cost of capital.* The Accounting Review 75, 13–41.

Bollen, K.A., 1990. *Overall fit in covariance structure models: Two types of sample size effects.* Psychological Bulletin 107, 256.

Bollen, K.A., 2002. *Latent variables in psychology and the social sciences.* Annual review of psychology 53, 605–634.

Bollen, K.A., 2014. *Structural equations with latent variables.* John Wiley & Sons.

Bollen, K.A., Lennox, R., 1991. *Conventional wisdom on measurement: A structural equation perspective.* Psychological bulletin 110, 305.

Bollen, K.A., Stine, R.A., 1992. *Bootstrapping goodness-of-fit measures in structural equation models.* Sociological Methods & Research 21, 205–229.

Bone, P.F., Sharma, S., Shimp, T.A., 1989. *A bootstrap procedure for evaluating goodness-of-fit indices of structural equation and confirmatory factor models.* Journal of Marketing Research 105–111.

Botosan, C.A., 1997. *Disclosure level and the cost of equity capital.* Accounting review 323–349.

Botosan, C.A., 2000. *Evidence that greater disclosure lowers the cost of equity capital.* Journal of applied corporate finance 12, 60–69.

Botosan, C.A., 2004. *Discussion of a framework for the analysis of firm risk communication.* The International Journal of Accounting 39, 289–295.

Botosan, C.A., Harris, M.S., 2000. *Motivations for a change in disclosure frequency and its consequences: An examination of voluntary quarterly segment disclosures.* Journal of Accounting Research 329–353.

Botosan, C.A., Plumlee, M.A., 2002. *A re-examination of disclosure level and the expected cost of equity capital.* Journal of accounting research 40, 21–40.

Bowen, R.M., Davis, A.K., Matsumoto, D.A., 2002. *Do conference calls affect analysts' forecasts?* The Accounting Review 77, 285–316.

Bowman, E.H., 1984. *Content analysis of annual reports for corporate strategy and risk*. Interfaces 14, 61–71.

Bozzolan, S., Mazzola, P., 2007. *Strategic plan presentations to financial analysts: the effects on earnings forecasts' revision and cost of capital,* in: Financial Reporting and Business Communication Conference, Cardiff Cerca Con Google.

Bozzolan, S., Trombetta, M., Beretta, S., 2009. *Forward-looking disclosures, financial verifiability and analysts' forecasts: A study of cross-listed European firms*. European Accounting Review 18, 435–473.

Bradshaw, M.T., 2011. *Analysts' forecasts: what do we know after decades of work?* Available at SSRN 1880339.

Breton, G., Taffler, R.J., 2001. *Accounting information and analyst stock recommendation decisions: a content analysis approach*. Accounting and business research 31, 91–101.

Brown, L.D., 1997. *Analyst forecasting errors: Additional evidence*. Financial Analysts Journal 53, 81–88.

Brown, L.D., Rozeff, M.S., 1978. *The superiority of analyst forecasts as measures of expectations: Evidence from earnings*. The Journal of Finance 33, 1–16.

Brown, S., Hillegeist, S.A., 2007. *How disclosure quality affects the level of information asymmetry*. Review of Accounting Studies 12, 443–477.

Brown, S., Hillegeist, S.A., Lo, K., 2004. *Conference calls and information asymmetry*. Journal of Accounting and Economics 37, 343–366.

Burrowes, A.W., Kastantin, J., Novicevic, M.M., 2004. *The Sarbanes-Oxley Act as a hologram of post-Enron disclosure: a critical realist commentary*. Critical Perspectives on Accounting 15, 797–811.

Bushee, B.J., Jung, M.J., Miller, G.S., 2011. *Conference presentations and the disclosure milieu.* Journal of Accounting Research 49, 1163–1192.

Bushee, B.J., Jung, M.J., Miller, G.S., 2013. *Do Investors Benefit from Selective Access to Management?* Available at SSRN 1880149.

Bushee, B.J., Leuz, C., 2005. *Economic consequences of SEC disclosure regulation: evidence from the OTC bulletin board*. Journal of accounting and economics 39, 233–264.

Bushee, B.J., Matsumoto, D.A., Miller, G.S., 2003. *Open versus closed conference calls: the determinants and effects of broadening access to disclosure*. Journal of Accounting and Economics 34, 149–180.

Bushee, B.J., Matsumoto, D.A., Miller, G.S., 2004. *Managerial and investor responses to disclosure regulation: The case of Reg FD and conference calls*. The Accounting Review 79, 617–643.

Bushee, B.J., Noe, C.F., 2000. *Corporate disclosure practices, institutional investors, and stock return volatility*. Journal of accounting research 171–202.

Bushman, R.M., Piotroski, J.D., Smith, A.J., 2004. *What determines corporate transparency?* Journal of accounting research 42, 207–252.

Buzby, S.L., 1974. *Selected items of information and their disclosure in annual reports*. Accounting Review 423–435.

Buzby, S.L., 1975. *Company size, listed versus unlisted stocks, and the extent of financial disclosure*. Journal of Accounting Research 13, 16–37.

Cabedo, J.D., Tirado, J.M., 2004. *The disclosure of risk in financial statements,* in: Accounting Forum. Elsevier, pp. 181–200.

Chang, M., D'Anna, G., Watson, I., Wee, M., 2008. *Does disclosure quality via investor relations affect information asymmetry?* Australian Journal of management 33, 375–390.

Cheng, C.A., Collins, D., Huang, H.H., 2006. *Shareholder rights, financial disclosure and the cost of equity capital.* Review of Quantitative Finance and Accounting 27, 175–204.

Cheng, E., Courtenay, S.M., 2006. *Board composition, regulatory regime and voluntary disclosure.* The International Journal of Accounting 41, 262–289.

Chen, S., Chen, X., Cheng, Q., 2008. *Do family firms provide more or less voluntary disclosure?* Journal of accounting research 46, 499–536.

Chen, S., DeFond, M.L., Park, C.W., 2002. *Voluntary disclosure of balance sheet information in quarterly earnings announcements.* Journal of Accounting and Economics 33, 229–251.

Chen, X., Cook, R.D., 2010. *Some insights into continuum regression and its asymptotic properties.* Biometrika 97, 985–989.

Chen, Z., Dhaliwal, D.S., Xie, H., 2010. *Regulation fair disclosure and the cost of equity capital.* Review of Accounting Studies 15, 106–144.

Chiyachantana, C.N., Jiang, C.X., Taechapiroontong, N., Wood, R., 2004. *The impact of Regulation Fair Disclosure on information asymmetry and trading: An intraday analysis*. The Financial Review 39, 549–577.

Choi, F.D., 1973. *Financial disclosure and entry to the European capital market.* Journal of Accounting Research 159–175.

Choi, F.D., 1974. *European Disclosure: The Competitive Disclosure Hypothesis*. Journal of International Business Studies 5, 15–23.

Clarkson, P.M., Kao, J.L., Richardson, G.D., 1999. *Evidence That Management Discussion and Analysis (MD&A) is a Part of a Firm's Overall Disclosure Package*. Contemporary Accounting Research 16, 111–134.

Clarkson, P.M., Kao, J.L., Richardson, G.D., 1994. *The Voluntary Inclusion of Forecasts in the MD&A Section of Annual Reports*. Contemporary Accounting Research 11, 423–450.

Cole, D.A., Maxwell, S.E., Arvey, R., Salas, E., 1993. *Multivariate group comparisons of variable systems: MANOVA and structural equation modeling*. Psychological Bulletin 114, 174.

Cooke, T.E., 1989a. *Disclosure in the corporate annual reports of Swedish companies. Accounting and business research* 19, 113–124.

Cooke, T.E., 1989b. *Voluntary corporate disclosure by Swedish companies*. Journal of International Financial Management & Accounting 1, 171–195.

Cooke, T.E., 1992. *The impact of size, stock market listing and industry type on disclosure in the annual reports of Japanese listed corporations*. Accounting and Business Research 22, 229–237.

Cooke, T.E., 1998. *Regression analysis in accounting disclosure studies*. Accounting and Business Research 28, 209–224.

Core, J.E., 2001. *A review of the empirical disclosure literature: discussion*. Journal of Accounting and Economics 31, 441–456.

Cormier, D., Magnan, M., Van Velthoven, B., 2005. *Environmental disclosure quality in large German companies: economic incentives, public pressures or institutional conditions?* European accounting review 14, 3–39.

Courtis, J.K., 1992. *The reliability of perception-based annual report disclosure studies.* Accounting and Business Research 23, 31–43.

Çürük, T., 2009. *An analysis of the companies' compliance with the EU disclosure requirements and corporate characteristics influencing it: A case study of Turkey*. Critical Perspectives on Accounting 20, 635–650.

Daske, H., Gebhardt, G., 2006. *International financial reporting standards and experts' perceptions of disclosure quality*. Abacus 42, 461–498.

Depoers, F., 2000. *A cost benefit study of voluntary disclosure: Some empirical evidence from French listed companies.* European Accounting Review 9, 245–263.

Dhaliwal, D.S., Khurana, I.K., Pereira, R., 2011. *Firm Disclosure Policy and the Choice Between Private and Public Debt.* Contemporary Accounting Research 28, 293–330.

Dhaliwal, D.S., Radhakrishnan, S., Tsang, A., Yong George Yang, 2012. *Nonfinancial Disclosure and Analyst Forecast Accuracy: International Evidence on Corporate Social Responsibility Disclosure.* Accounting Review 87, 723–759. doi:10.2308/accr-10218

Diamantopoulos, A., 1994. *Modelling with LISREL: A guide for the uninitiated*. Journal of Marketing Management 10, 105–136.

Dobler, M., Lajili, K., Zéghal, D., 2011. *Attributes of corporate risk disclosure: An international investigation in the manufacturing sector*. Journal of International Accounting Research 10, 1–22.

Drake, M.S., Myers, J.N., Myers, L.A., 2009. *Disclosure quality and the mispricing of accruals and cash flow*. Journal of Accounting, Auditing & Finance 24, 357–384.

Duarte, J., Han, X., Harford, J., Young, L., 2008. *Information asymmetry, information dissemination and the effect of regulation FD on the cost of capital*. Journal of Financial Economics 87, 24–44.

Dutta, S., Trueman, B., 2002. *The interpretation of information and corporate disclosure strategies*. Review of Accounting Studies 7, 75–96.

Dye, R.A., 1986. *Proprietary and nonproprietary disclosures*. Journal of business 331–366.

Eleswarapu, V.R., Thompson, R., Venkataraman, K., 2004. *The impact of Regulation Fair Disclosure: Trading costs and information asymmetry*. Journal of Financial and Quantitative Analysis 39, 209–225.

Elliott, R.K., Jacobson, P.D., 1994. *Costs and Benefits of Business information*. Accounting Horizons 8, 80–96.

Enders, C.K., 2001. *The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data*. Psychological methods 6, 352.

Ertimur, Y., 2007. *Discussion of "How disclosure quality affects the level of information asymmetry."* Review of Accounting Studies 12, 479–485.

Fabrigar, L.R., Porter, R.D., Norris, M.E., 2010. *Some things you should know about structural equation modeling but never thought to ask.* Journal of Consumer Psychology 20, 221–225. doi:10.1016/j.jcps.2010.03.003

Farvaque, E., Refait-Alexandre, C., Saïdane, D., 2011. *Corporate disclosure: a review of its (direct and indirect) benefits and costs*. International Economics 128, 5–31. doi:10.1016/S2110-7017(13)60001-3

Fayers, P.M., Hand, D.J., 2002. *Causal variables, indicator variables and measurement scales: an example from quality of life*. Journal of the Royal Statistical Society: Series A (Statistics in Society) 165, 233–253.

Field, L., Lowry, M., Shu, S., 2005. *Does disclosure deter or trigger litigation?* Journal of Accounting and Economics 39, 487–507.

Firth, M., 1979. *The impact of size, stock market listing, and auditors on voluntary disclosure in corporate annual reports*. Accounting and Business Research 9, 273–280.

Fishman, M.J., Hagerty, K.M., 1989. *Disclosure decisions by firms and the competition for price efficiency*. The Journal of Finance 44, 633–646.

Fogarty, T.J., Rogers, R.K., 2005. *Financial analysts' reports: an extended institutional theory evaluation*. Accounting, Organizations and Society 30, 331–356.

Forker, J.J., 1984. *Contract Value Accounting and the Monitoring of Managerial Performance: An Agency-Based Proposal*. Accounting and Business Research 14, 125–137.

Fornell, C., Larcker, D.F., 1981a. *Evaluating structural equation models with unobservable variables and measurement error*. Journal of marketing research 39–50.

Fornell, C., Larcker, D.F., 1981b. *Structural equation models with unobservable variables and measurement error: Algebra and statistics*. Journal of marketing research 382–388.

Francis, J., Nanda, D., Wang, X., 2006. *Re-examining the effects of regulation fair disclosure using foreign listed firms to control for concurrent shocks*. Journal of Accounting and Economics 41, 271–292.

Francis, J.R., Khurana, I.K., Pereira, R., 2005. *Disclosure Incentives and Effects on Cost of Capital around the World*. The Accounting Review 80, 1125–1162.

Frankel, R., Johnson, M., Skinner, D.J., 1999. *An empirical examination of conference calls as a voluntary disclosure medium.* Journal of Accounting Research 133–150.

Frankel, R., McNichols, M., Wilson, G.P., 1995. *Discretionary disclosure and external financing*. Accounting Review 135–150.

Gao, S., Mokhtarian, P.L., Johnston, R.A., 2008. *Nonnormality of data in structural equation models.* Transportation Research Record: Journal of the Transportation Research Board 2082, 116–124.

Gelb, D.S., Strawser, J.A., 2001. *Corporate social responsibility and financial disclosures: an alternative explanation for increased disclosure*. Journal of Business Ethics 33, 1–13.

Gibbins, M., Richardson, A., Waterhouse, J., 1990. *The management of corporate financial disclosure: opportunism, ritualism, policies, and processes*. Journal of accounting research 121–143.

Gigler, F., 1994. *Self-enforcing voluntary disclosures.* Journal of Accounting Research 224–240.

Gintschel, A., Markov, S., 2004. *The effectiveness of Regulation FD.* Journal of Accounting and Economics 37, 293–314.

Glover, J.C., 2012. *Disclosure and incentives*. Accounting Horizons 26, 371–380.

Goffin, R.D., 2007. *Assessing the adequacy of structural equation models: Golden rules and editorial policies.* Personality and Individual Differences 42, 831–839.

Goldberger, A.S., 1972. *Structural equation methods in the social sciences.* Econometrica: Journal of the Econometric Society 979–1001.

Gonedes, N.J., Dopuch, N., Penman, S.H., 1976. *Disclosure rules, information-production, and capital market equilibrium: The case of forecast disclosure rules*. Journal of Accounting Research 89–137.

Gray, S.J., 1988. *Towards a theory of cultural influence on the development of accounting systems internationally.* Abacus 24, 1–15.

Gregson, T., 1992. *The advantages of LISREL for accounting researchers*. American Journal of Sociology 1, 16.

Grüning, M., 2007. *Drivers of corporate disclosure: a structural equation analysis in a Central European setting.* Management Research News 30, 646–660.

Grüning, M., 2011. *Artificial intelligence measurement of disclosure (AIMD).* European Accounting Review 20, 485–519.

Guadagnoli, E., Velicer, W.F., 1988. *Relation to sample size to the stability of component patterns.* Psychological bulletin 103, 265.

Gumb, B., 2007. *What is shown, what is hidden: Compulsory disclosure as a spectacle*. Critical Perspectives on Accounting 18, 807–828.

Guthrie, J., Parker, L.D., 1989. *Corporate social reporting: a rebuttal of legitimacy theory.* Accounting and business research 19, 343–352.

Haavelmo, T., 1943. *The statistical implications of a system of simultaneous equations*. Econometrica, Journal of the Econometric Society 1–12.

Hassan, O., Marston, C., 2010. *Disclosure measurement in the empirical accounting literature-a review article.* Available at SSRN 1640598.

Hasseldine, J., Salama, A.I., Toms, J.S., 2005. *Quantity versus quality: the impact of environmental disclosures on the reputations of UK Plcs*. The British Accounting Review 37, 231–248.

Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., Boulianne, S., 2007. *Testing! testing! one, two, three–Testing the theory in structural equation models!* Personality and Individual Differences 42, 841–850.

Healy, P.M., Hutton, A.P., Palepu, K.G., 1999. *Stock performance and intermediation changes surrounding sustained increases in disclosure.* Contemporary accounting research 16, 485–520.

Healy, P.M., Palepu, K.G., 1993. *The effect of firms' financial disclosure strategies on stock prices.* Accounting Horizons 7, 1.

Healy, P.M., Palepu, K.G., 2001. *Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature.* Journal of accounting and economics 31, 405–440.

Heflin, F., Hsu, C., 2008. *The impact of the SEC's regulation of non-GAAP disclosures.* Journal of Accounting and Economics 46, 349–365.

Heflin, F., Subramanyam, K.R., Zhang, Y., 2003. *Regulation FD and the financial information environment: Early evidence.* The Accounting Review 78, 1–37.

Heitzman, S., Wasley, C., Zimmerman, J., 2010. *The joint effects of materiality thresholds and voluntary disclosure incentives on firms' disclosure decisions*. Journal of accounting and economics 49, 109–132.

Hodgdon, C., Tondkar, R.H., Harless, D.W., Adhikari, A., 2008. *Compliance with IFRS disclosure requirements and individual analysts' forecast errors*. Journal of International Accounting, Auditing and Taxation 17, 1–13.

Hofstede, G., 1984. *Culture's consequences: International differences in work-related values* (Vol. 5). Sage.

Holland, J.B., 1998. *Private disclosure and financial reporting*. Accounting and Business Research 28, 255–269.

Holland, J.B., 2005. *A grounded theory of corporate disclosure*. Accounting and business research 35, 249–267.

Hope, O.K., 2003. *Disclosure practices, enforcement of accounting standards, and analysts' forecast accuracy: An international study*. Journal of accounting research 41, 235–272.

Hussainey, K., Schleicher, T., Walker, M., 2003. *Undertaking large-scale disclosure studies when AIMR-FAF ratings are not available: the case of prices leading earnings*. Accounting and Business Research 33, 275–294.

Iacobucci, D., 2009. *Everything you always wanted to know about SEM (structural equations modeling) but were afraid to ask.* Journal of Consumer Psychology 19, 673–680. doi:10.1016/j.jcps.2009.09.002

Iacobucci, D., 2010. *Structural equations modeling: Fit Indices, sample size, and advanced topics*. Journal of Consumer Psychology 20, 90–98. doi:10.1016/j.jcps.2009.09.003

ICAEW, 1999. *No Surprises: The Case for Better Risk Reporting.*

ICAEW, 2002. *Reporting Business Risks: Meeting Expectations.*

ICAEW, 2004. *Information for Better Markets.*

ICAEW, 2009. *Developments in New Reporting Models.*

ICAEW, 2011. *No Surprises: Working for Better Risk Reporting.*

Irani, A.J., Karamanou, I., 2003. *Regulation fair disclosure, analyst following, and analyst forecast dispersion.* Accounting Horizons 17, 15–29.

James, L.R., Singh, B.K., 1978. *An introduction to the logic, assumptions, and basic analytic procedures of two-stage least squares*. Psychological Bulletin 85, 1104.

Jarvis, C.B., MacKenzie, S.B., Podsakoff, P.M., 2003. *A critical review of construct indicators and measurement model misspecification in marketing and consumer research*. Journal of consumer research 30, 199–218.

Jensen, M.C., Meckling, W.H., 1976. *Theory of the firm: Managerial behavior, agency costs and ownership structure.* Journal of financial economics 3, 305–360.

Jiang, H., Habib, A., Hu, B., 2011. *Ownership concentration, voluntary disclosures and information asymmetry in New Zealand.* The British Accounting Review 43, 39–53.

Johnson, R.A., Greening, D.W., 1999. *The effects of corporate governance and institutional ownership types on corporate social performance*. Academy of Management Journal 42, 564–576.

Jöreskog, K.G., Sörbom, D., 1982. *Recent developments in structural equation modeling.* Journal of Marketing Research 404–416.

Jorion, P., Liu, Z., Shi, C., 2005. *Informational effects of regulation FD: evidence from rating agencies.* Journal of financial economics 76, 309–330.

Kenny, D.A., Judd, C.M., 1984. *Estimating the nonlinear and interactive effects of latent variables.* Psychological bulletin 96, 201.

Khurana, I.K., Pereira, R., Martin, X., 2006. *Firm Growth and Disclosure: An Empirical Analysis.* The Journal of Financial and Quantitative Analysis 41, 357–380.

Kim, O., 1993. *Disagreements among shareholders over a firm's disclosure policy*. The Journal of Finance 48, 747–760.

Kim, O., Verrecchia, R.E., 2001. *The relation among disclosure, returns, and trading volume information.* The Accounting Review 76, 633–654.

Kinney, J.R., William, R., Shepardson, M.L., 2011. *Do control effectiveness disclosures require SOX 404 (b) internal control audits? A natural experiment with small US public companies.* Journal of Accounting Research 49, 413–448.

Klein, L.R., 1960. *Single equation vs. equation system methods of estimation in econometrics.* Econometrica: Journal of the Econometric Society 866–871.

Kothari, S.P., Li, X., Short, J.E., 2009. *The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis.* The Accounting Review 84, 1639–1670.

Kothari, S.P., Shu, S., Wysocki, P.D., 2009. *Do managers withhold bad news?* Journal of Accounting Research 47, 241–276.

Kroll, M., Wright, P., Heiens, R.A., 1999. *The contribution of product quality to competitive advantage: impacts on systematic variance and unexplained variance in returns*. Strategic Management Journal 20, 375–384.

Kross, W.J., Suk, I., 2012. *Does Regulation FD work? Evidence from analysts' reliance on public disclosure.* Journal of Accounting and Economics 53, 225–248.

Kupek, E., 2005. *Log-Linear Transformation of Binary Variables: A Suitable Input for SEM.* Structural Equation Modeling: A Multidisciplinary Journal 12, 28–40. doi:10.1207/s15328007sem1201_2

Lajili, K., Zéghal, D., 2005. *A content analysis of risk management disclosures in Canadian annual reports.* Canadian Journal of Administrative Sciences 22, 125.

Lambert, R., Leuz, C., Verrecchia, R.E., 2007. *Accounting information, disclosure, and the cost of capital.* Journal of accounting research 45, 385–420.

Lang, M.H, Lins, K.V., Maffett, M., 2012. *Transparency, liquidity, and valuation: International evidence on when transparency matters most.* Journal of Accounting Research 50, 729–774.

Lang, M.H., Lins, K.V., Miller, G.S., 2003. *ADRs, Analysts, and Accuracy: Does Cross Listing in the United States Improve a Firm's Information Environment and Increase Market Value?* Journal of Accounting Research 41, 317–345.

Lang, M., Lundholm, R., 1993. *Cross-sectional determinants of analyst ratings of corporate disclosures.* Journal of accounting research 246–271.

Lang, M., Lundholm, R.J., 1996a. *The Relation Between Security Returns, Firm Earnings, and Industry Earnings.* Contemporary Accounting Research 13, 607–629.

Lang, M.H., Lundholm, R.J., 1996b. *Corporate disclosure policy and analyst behavior*. Accounting review 467–492.

Lang, M.H., Lundholm, R.J., 2000. *Voluntary Disclosure and Equity Offerings: Reducing Information Asymmetry or Hyping the Stock?.* Contemporary accounting research 17, 623–662.

Lev, B., Penman, S.H., 1990. *Voluntary forecast disclosure, nondisclosure, and stock prices. Journal of Accounting Research* 49–76.

Lewellen, W.G., Park, T., Ro, B.T., 1996. *Self-serving behavior in managers' discretionary information disclosure decisions.* Journal of accounting and Economics 21, 227–251.

Li, L., Dennis Cook, R., Nachtsheim, C.J., 2005. *Model-free variable selection.* Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 285–299.

Linsley, P.M., Lawrence, M.J., 2007. *Risk reporting by the largest UK companies: readability and lack of obfuscation.* Accounting, Auditing & Accountability Journal 20, 620–627.

Linsley, P.M., Shrives, P.J., 2006. *Risk reporting: A study of risk disclosures in the annual reports of UK companies.* The British Accounting Review 38, 387–404.

Linsley, P.M., Shrives, P.J., Crumpton, M., 2006. *Risk disclosure: An exploratory study of UK and Canadian banks.* Journal of Banking Regulation 7, 268–282.

Lopes, P.T., Rodrigues, L.L., 2007. *Accounting for financial instruments: An analysis of the determinants of disclosure in the Portuguese stock exchange.* The International Journal of Accounting 42, 25–56.

Lundholm, R., Myers, L.A., 2002. *Bringing the future forward: the effect of disclosure on the returns-earnings relation.* Journal of Accounting Research 40, 809–839.

Lundholm, R., Van Winkle, M., 2006. *Motives for disclosure and non-disclosure: a framework and review of the evidence.* Accounting and Business Research 36, 43–48.

MacKenzie, S.B., Podsakoff, P.M., Jarvis, C.B., 2005. *The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions.* Journal of Applied Psychology 90, 710.

Mainul Islam, M.D., Faniran, O.O., 2005. *Structural equation model of project planning effectiveness.* Construction Management and Economics 23, 215–223.

Malone, D., Fries, C., Jones, T., 1993. *An empirical investigation of the extent of corporate financial disclosure in the oil and gas industry*. Journal of Accounting, Auditing & Finance 8, 249–273.

Marshall, A.P., Weetman, P., 2002. *Information asymmetry in disclosure of foreign exchange risk management: can regulation be effective?* Journal of Economics and Business 54, 31–53.

Marshall, A.P, Weetman, P., 2007. *Modelling transparency in disclosure: the case of foreign exchange risk management.* Journal of Business Finance & Accounting 34, 705–739.

Marsh, H.W., Balla, J.R., McDonald, R.P., 1988. *Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size.* Psychological bulletin 103, 391.

Marston, C.L., Shrives, P.J., 1991. *The use of disclosure indices in accounting research: a review article.* The British Accounting Review 23, 195–210.

Mayew, W.J., 2008. *Evidence of management discrimination among analysts during earnings conference calls.* Journal of Accounting Research 46, 627–659.

McDonald, R.P., Ho, M.H.R., 2002. *Principles and practice in reporting structural equation analyses.* Psychological methods 7, 64.

McDonald, R.P., Marsh, H.W., 1990. *Choosing a multivariate model: Noncentrality and goodness of fit.* Psychological bulletin 107, 247.

Meek, G.K., Roberts, C.B., Gray, S.J., 1995. *Factors influencing voluntary annual report disclosures by US, UK and continental European multinational corporations*. Journal of international business studies 555–572.

Miles, J., Shevlin, M., 2007. *A time and a place for incremental fit indices*. Personality and Individual Differences 42, 869–874.

Millsap, R.E., 2007. *Structural equation modeling made difficult.* Personality and Individual Differences 42, 875–881.

Milne, M.J., 2002. *Positive accounting theory, political costs and social disclosure analyses: A critical look.* Critical perspectives on accounting 13, 369–395.

Moore, M.L., Buzby, S., 1972. *The quality of corporate financial disclosure: A comment.* Accounting review 581–584.

Morris, R.D., 1987. *Signalling, agency theory and accounting policy choice.* Accounting and Business Research 18, 47–56.

Mulaik, S., 2007. *There is a place for approximate fit in structural equation modelling.* Personality and Individual Differences 42, 883–891.

Mulaik, S.A., James, L.R., Van Alstine, J., Bennett, N., Lind, S., Stilwell, C.D., 1989. *Evaluation of goodness-of-fit indices for structural equation models.* Psychological bulletin 105, 430.

Muller, P., Pénin, J., 2006. *Why do firms disclose knowledge and how does it matter?* Journal of Evolutionary Economics 16, 85–108.

Orens, R., Aerts, W., Lybaert, N., 2009. *Intellectual capital disclosure, cost of finance and firm value.* Management Decision 47, 1536–1554.

Patelli, L., Prencipe, A., 2007. *The relationship between voluntary disclosure and independent directors in the presence of a dominant shareholder.* European Accounting Review 16, 5–33.

Patten, D.M., 1992. *Intra-industry environmental disclosures in response to the Alaskan oil spill: a note on legitimacy theory.* Accounting, Organizations and Society 17, 471–475.

Penman, S.H., 1980. *An empirical investigation of the voluntary disclosure of corporate earnings forecasts.* Journal of accounting research 132–160.

Petter, S., Straub, D., Rai, A., 2007. *Specifying formative constructs in information systems research.* MIS Quarterly 623–656.

Pownall, G., Wasley, C., Waymire, G., 1993. *The stock price effects of alternative types of management earnings forecasts.* Accounting Review 896–912.

Poznanski, P.J., Bline, D., 1997. *Using structural equation modeling to investigate the causal ordering of job satisfaction and organizational commitment among staff accountants*. Behavioral Research in Accounting 9, 154–171.

Preacher, K.J., Merkle, E.C., 2012. *The problem of model selection uncertainty in structural equation modeling.* Psychological methods 17, 1–14.

Raffournier, B., 1995. *The determinants of voluntary financial disclosure by Swiss listed companies.* European Accounting Review 4, 261–280.

Robb, S.W., Zarzeski, M.T., Single, L.E., 2001. *Nonfinancial disclosures across Anglo-American countries.* Journal of International Accounting, Auditing and Taxation 10, 71–83.

Roberts, J., Sanderson, P., Barker, R., Hendry, J., 2006. *In the mirror of the market: The disciplinary effects of company/fund manager meetings.* Accounting, Organizations and Society 31, 277–294.

Rogers, J.L., Van Buskirk, A., 2009. *Shareholder litigation and changes in disclosure behavior.* Journal of Accounting and Economics 47, 136–156.

Rowley, T., Berman, S., 2000. *A brand new brand of corporate social performance*. Business & society 39, 397–418.

Ryan, P., 2005. *The market impact of directors' trades: relationship to various measures of a firm's information environment.* The British Accounting Review 37, 319–337.

Ryan, P., Taffler, R.J., 2004. *Are Economically Significant Stock Returns and Trading Volumes Driven by Firm-specific News Releases?* Journal of Business Finance & Accounting 31, 49–82.

Saudagaran, S.M., Biddle, G.C., 1995. *Foreign listing location: A study of MNCs and stock exchanges in eight countries.* Journal of International Business Studies 26, 319–341.

Schadewitz, H.J., Kanto, A.J., 2002. *The impact of disclosure on the market response to reported earnings.* Scandinavian Journal of Management 18, 521–542.

Schleicher, T., Hussainey, K., Walker, M., 2007. *Loss firms' annual report narratives and share price anticipation of earnings.* The British Accounting Review 39, 153–171.

Schrand, C.M., Elliott, J.A., 1998. *Risk and financial reporting: A summary of the discussion at the 1997 AAA/FASB conference.* Accounting Horizons 12, 271–282.

Sengupta, P., 1998. *Corporate Disclosure Quality and the Cost of Debt*. The Accounting Review 73, 459–474.

Sethi, S.P., 1979. *A conceptual framework for environmental analysis of social issues and evaluation of business response patterns.* Academy of Management Review 4, 63–74.

Sharma, S., Durvasula, S., Dillon, W.R., 1989. *Some results on the behavior of alternate covariance structure estimation procedures in the presence of non-normal data.* Journal of Marketing research XXVI, 214–221.

Shocker, A.D., Sethi, S.P., 1973. *An Approach to Incorporating Societal Preferences in Developing Corporate Action Strategies.* California Management Review 15, 97–105. doi:10.2307/41164466

Shook, C.L., Ketchen, D.J., Hult, G.T.M., Kacmar, K.M., 2004. *An assessment of the use of structural equation modeling in strategic management research.* Strategic management journal 25, 397–404.

Singhvi, S.S., Desai, H.B., 1971. *An empirical analysis of the quality of corporate financial disclosure.* Accounting review 46, 129–138.

Skinner, D.J., 1994. *Why firms voluntarily disclose bad news.* Journal of accounting research 32, 38–60.

Slack, R., Campbell, D., 2008. *Narrative reporting: analysts' perceptions of its value and relevance.* Association of Chartered Certified Accountants.

Smith, D., Langfield-Smith, K., 2004. *Structural equation modeling in management accounting research: critical analysis and opportunities.* Journal of Accounting Literature 23, 49–86.

Spence, M., 1973. *Job market signaling*. The quarterly journal of Economics 87, 355–374.

Stanton, P., Stanton, J., 2002. *Corporate annual reports: research perspectives used*. Accounting, Auditing & Accountability Journal 15, 478–500.

Steenkamp, J.-B.E., Baumgartner, H., 2000. *On the use of structural equation models for marketing modeling.* International Journal of Research in Marketing 17, 195–202.

Steiger, J.H., 2007. *Understanding the limitations of global fit assessment in structural equation modeling.* Personality and Individual differences 42, 893–898.

Suijs, J., 2007. *Voluntary disclosure of information when firms are uncertain of investor response.* Journal of Accounting and Economics 43, 391–410.

Taylor, G., Tower, G., Neilson, J., 2010. *Corporate communication of financial risk.* Accounting & Finance 50, 417–446.

Toms, J.S., 2002. *Firm resources, quality signals and the determinants of corporate environmental reputation: some UK evidence.* The British Accounting Review 34, 257–282.

van Staden, C.J., Hooks, J., 2007. *A comprehensive comparison of corporate environmental reporting and responsiveness*. The British Accounting Review 39, 197–210.

Vanstraelen, A., Zarzeski, M.T., Robb, S.W., 2003. *Corporate nonfinancial disclosure practices and financial analyst forecast ability across three European countries*. Journal of International Financial Management & Accounting 14, 249–278.

Veld, C., Veld-Merkoulova, Y.V., 2008. *The risk perceptions of individual investors*. Journal of Economic Psychology 29, 226–252.

Verrecchia, R.E., 2001. *Essays on disclosure*. Journal of accounting and economics 32, 97–180.

Walker, M., Louvari, E., 2003. *The determinants of voluntary disclosure of adjusted earnings per share measures by UK quoted companies.* Accounting and Business Research 33, 295–309.

Watson, A., Shrives, P., Marston, C., 2002. *Voluntary disclosure of accounting ratios in the UK.* The British Accounting Review 34, 289–313.

Watts, R.L., Zimmerman, J.L., 1978. *Towards a positive theory of the determination of accounting standards.* Accounting review 53, 112–134.

Watts, R.L., Zimmerman, J.L., 1979. *The demand for and supply of accounting theories: the market for excuses.* Accounting Review 54, 273–305.

Watts, R.L., Zimmerman, J.L., 1990. *Positive accounting theory: a ten year perspective.* Accounting review 65, 131–156.

Welker, M., 1995. *Disclosure Policy, Information Asymmetry, and Liquidity in Equity Markets.* Contemporary accounting research 11, 801–827.

Widaman, K.F., Thompson, J.S., 2003. *On specifying the null model for incremental fit indices in structural equation modeling.* Psychological methods 8, 16–37.

Wright, S., 1960. *Path coefficients and path regressions: alternative or complementary concepts?* Biometrics 16, 189–202.

Wright, S., 1921. *Correlation and causation*. Journal of agricultural research 20, 557–585.

Yuan, K.-H., Chan, W., Bentler, P.M., 2000. *Robust transformation with applications to structural equation modelling*. British Journal of Mathematical and Statistical Psychology 53, 31–50.

Zarzeski, M.T., 1996. *Spontaneous harmonization effects of culture and market forces on accounting disclosures practices*. Accounting Horizons 10, 18–37.